UNIVERSITY OF COPENHAGEN

FACULTY OF HEALTH AND MEDICAL SCIENCES

# Automatic learning and pattern recognition using sensor data in livestock farming

PhD Thesis · Dan Børge Jensen · 2016

# Automatic learning and pattern recognition using sensor data in livestock farming

PhD Thesis

Dan Børge Jensen

HERD

Centre for Herd-oriented Education, Research and Development

Department of Large Animal Sciences

University of Copenhagen

February 2016

**Principal advisors**

Professor Nils Toft (March 2013 - August 2013)
Department of Large Animal Sciences
University of Copenhagen, Denmark

Associate professor Cécile Cornou (August 2013 - October 2014)
Department of Large Animal Sciences
University of Copenhagen, Denmark

Professor Anders Ringgaard Kristensen (October 2014 - February 2016)
Department of Large Animal Sciences
University of Copenhagen, Denmark

**Co-advisors**

Associate professor Cécile Cornou (August 2013 - October 2014)
Department of Large Animal Sciences
University of Copenhagen, Denmark

Professor Nils Toft (August 2013 - February 2016)
Section for Epidemiology
Technical University of Denmark, Denmark

**External advisor and contact person while staying abroad**

Associate professor Albert De Vries
Department of Animal Sciences
University of Florida, United States of America

**Assessment committee**

Associate Professor Matt Denwood (Chairman)
Department of Large Animal Sciences
University of Copenhagen, Denmark

Dr Søren Lundbye-Christensen
Unit of Clinical Biostatistics and Bioinformatics
Aalborg University Hospital, Denmark

Associate Professor Lluis Miguel Plà Aragones
Department of Mathematics
University of Lleida, Spain

*This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen 29/02/2016.*

"We wish to find the truth, no matter where it lies. But to find the truth we need imagination and skepticism both. We will not be afraid to speculate, but we will be careful to distinguish speculation from fact."

— Carl Sagan

## Preface and acknowledgements

When I accepted the PhD position at the Department of Large Animal Sciences at the University of Copenhagen, it marked a series of important changes in my life. First of all, it meant leaving the familiar surroundings of the DTU campus in Lyngby. Of course after six years of living, studying, and working within that campus, a change of scenery was refreshing to say the least. This change was made even more appealing by the beautiful surroundings at the Frederiksberg campus where my office would be, as well as the friendly people who would become my colleagues.

It was also a significant change of a more academic nature; I remember being asked in my job interview if I would be able to work with "such people as veterinarians". My immediate response was something along the lines of a Pippi Longstocking-esk "I have never done that before, so I don't see why not". At the time, I saw no reason to think that veterinarians (or animal scientists who, to be fair, made up the majority of my actual colleagues and collaboration partners) would be any different from the collection of mostly engineers, with whom I had worked and associated until that point. I should however soon learn that my new colleagues had a shared frame of reference, which was different from my own. This difference would at times make their reasoning, or even casual conversations, somewhat tricky to follow. Far from being a problem, however, it is my experience and humble belief that differing minds can lead to the most interesting conversations as well as the most fruitful collaborations, and I have thoroughly enjoyed many interesting conversations with my fellow PhD students at the Frederiksberg campus. Because of this, I feel that the change from DTU to the Department of Large Animal Sciences has been a most enriching experience.

As another example of the value of knowing a diverse multitude of people, my old friend Nicole Egesby deserves a special mentioning. She recently passed her farmers exam after training in a commercial Danish pig farm for a total of three years. When I started writing my introduction chapter to this thesis, I was struck by writers block. Her stories of her daily routines and practical experiences from the farm helped me to get the writing flowing, and in the end I ended up with what I consider to be a very decent introduction to my thesis. For this, I wish to thank her.

As my readers can probably guess from the lists of advisors in the beginning of this thesis, I have seen my fair share of turnover in principal advisors during my three years as a PhD student. This has admittedly been somewhat stressful at times, but it is also worth acknowledging the fact that each of my three main advisors brought value to my project in their own way. Interestingly enough, in retrospect they all seem to have been the exact advisors I needed at their respective times.

From Nils Toft, my first main advisor, I could always count on brutally honest and unapologetic criticism of my work. This may sound like a harsh treatment, and sometimes it even felt that way, but all of my work, which has been subjected to his criticism, has been considerably improved by it. Given the choice, I would not have been without it, and I owe him many thanks.

Cécile Cornou, my second main advisor, excelled at boosting my confidence at times when I was unsure of myself and unsure of whether I would be able to finish this PhD in any meaningful way, as I think all PhD students go through in their early to mid phases of their projects. Moreover, she seemed to know exactly when to push me forward on the right track when I would gone in a different direction if left to my own devices. She also knew when to slow me down and make sure that I had the basics properly covered, when I wanted to bite off more than I could have handled at the time. I am very thankful for having received her guidance when I did.

Ander Ringgaard Kristensen, my third and last main advisor who advised me for the last year or so of my PhD project, generally demonstrated a confidence that I was able to independently pursue my research without too much interference. Even so, he would always set aside plenty of time to meet with me when I felt it to be necessary, sometimes with as little as an hour's notice. It was while having him as my advisor that I began to truly believe that I could make real and meaningful contributions to the fields of animal science and specifically precision livestock science. I thank him for his help and advices, and for the opportunity to continue making my scientific contributions as his post doc after earning my PhD degree.

Even though Albert De Vries was never officially my advisor, he was nonetheless the man who served as my contact person and primary source of advice during the half year or so when I stayed in Gainesville, Florida, as well as the two weeks when I stayed at Wageningen University in the Netherlands. More than that, Albert invited me into his home to have dinner with his lovely family on numerous occasions. For all of this, he deserves a special thank you.

When mentioning Wageningen University, I should really also mention Henk Hogeveen, who deserves to be thanked for letting me have the opportunity to come and work in his department at said university. In addition to a mere office space, Henk offered interesting conversations on the subject of detection systems. These conversations have greatly influenced several papers presented in this thesis, and I imagine they will continue to do for the remainder of my research career.

Of course none of this work would have been possible without the funding needed to do it. For this, I wish to thank the Danish Council for Strategic Research (Grant number 11-116191).

On the home front, I think my dear girlfriend Majbritt Toldbod Nielsen deserves a thank you for putting up with me during the more stressful parts of my writing process. During these times, I could apparently and often unbeknownst to myself be rather unbearable to be around. Similarly, my parents deserves my gratitude for offering me a work station in their house in Lumsås, whenever I just needed to get away from either home or the office to do my work.

On that note, I think my father Jan Børge Jensen deserves a very special and thankful mentioning. As far back as I can remember, he would inspire me to feel awe and wonderment of the world around us, and of science as way of understanding and utilizing this world. I remember how he explained the emergence of the seasons using an orange and my bedroom lamp, how he explained that the stars in the night sky were so distant that what we saw were actually their distant past, and how he taught me to make working parachutes for my toys. And many more examples could be given. In short, he led me on the path towards science with no formal scientific training of his own, and I would not be anywhere near where I am today, had it not been for him.

# Summary

This thesis should be considered in the lights of two primary contexts, namely the project context and the societal context.

In terms of the project context, the PhD project, which formed the basis for this thesis, was itself a part of a larger project called *PigIT - improving welfare and productivity in growing pigs using advanced ICT methods*. This overall project covered several sub-projects, each of which were related to a number of different work packages with different foci. The continuing overall goal of this larger project is, in essence, to simultaneously improve both animal welfare and productivity of growing pigs. This is done by implementing existing sensor technologies for collecting precision data for monitoring purposes. By precision data is meant regularly and preferably automatically collected data pertaining to specific animals within the herd. For the PhD project presented in this thesis, the focus was on combining the collected data and, by means of models and classification tools, be able to predict or detect undesired events based on these combined data.

In terms of the societal context, animal welfare is known to be a great concern for the citizens of the European Union, and in particular in the Scandinavian countries, including Denmark. It is however also known that the average European consumer is not willing to pay extra for animal products made under above-standard welfare conditions. If the productivity can be improved by improving the welfare, this would become less of a problem. Another important societal aspect is the rise of antibiotic resistant bacteria, which are naturally evolving in response to antibiotics use, including the antibiotics use seen in modern animal farming. It seems reasonable to assume that improving the health of the pigs would result in a reduced need for using antibiotics, which in return would reduce the selective pressure for antibiotic resistance in bacteria, thus slowing this evolutionary process down.

Two primary hypotheses have been the motivating premises behind the research presented in this thesis. These are referred to as the *environment hypothesis* and the *normality hypothesis*. In short, the environment hypothesis states that the environment experienced by the pigs, including the very local environment at *e.g.* the pen level, affects the health and comfort of the pigs. Thus monitoring the local environment could yield predictions of undesired events. The normality hypothesis states that so long as the animals are healthy and comfortable, their behavior and physiological characteristics will be predictable by a dynamic model designed to describe this normal state. It then follows that when the animals become sick or experience discomfort, the same model will fail in predicting the behavior and physiological characteristics of the same animals. Thus the inaccuracies in the forecasts of such a model could be used to provide early warnings of oncoming problems.

The five papers discussed in this thesis take one or both of the above mentioned hypotheses as their premises. Together they show a cumulative progression towards reaching the goals of the PhD project presented in this thesis, namely to combine data from multiple sources for the purpose of predicting or detecting undesired events in growing pigs:

**Paper 1** demonstrates that there is good reason to suspect that monitoring the pen level environment will provide pen specific information on the health and welfare of the animals. It should be noted that since no direct pen specific data or direct health and welfare related registrations were available, proxies had to be used.

**Paper 2** shows that daily summaries of automatically collected pen level temperatures can provide information which is directly useful for predicting the onset of diarrhea and pen fouling at the pen level.

**Paper 3** shows that very diverse data streams can be meaningfully combined using a multivariate dynamic linear model (DLM). By diverse data streams is meant data produced by different sensors pertaining to different variables (specifically water consumption, feed amount, and live weight) with differing numerical values and variances, and different observational frequencies. Paper 3 further presents a method for unifying the multiple forecast errors made by the multivariate DLM at each observation step. This unification method is based on Cholesky decomposition. The combination of a multivariate DLM for modeling multiple data streams and the Cholesky-based unification of the forecast errors is called the *DLM/Cholesky method*.

**Paper 4** shows that the DLM/Cholesky method can be used to make indiscriminant pen level predictions of undesired events in a large scale data set collected in a commercial Danish pig farm. It should be noted that only registrations of diarrhea and pen fouling were available to evaluate the method's performance. Furthermore, Paper 4 demonstrates that monitoring the data streams via multivariate DLMs provide a simple method of estimating the relative information value of the various data streams. This is achieved by systematically including or omitting specific data streams while estimating the resulting performance. Lastly, Paper 4 showed that the information value of the pen level temperature data was much lower when monitored with a multivariate DLM compared to the summary method used in Paper 2.

**Paper 5** demonstrated how a naïve Bayesian classifier (NBC) could be used to combine the forecast errors from a multivariate DLM as well as categorical non-sensor data for the purpose of detecting undesired events. This method was used to detect mastitis in dairy cows from Florida.

These five papers all show that precision data are useful and important for detecting or predicting undesired events in groups or individual animals. Together, they further show that the multivariate DLM is a useful approach for monitoring the animal-oriented data, but that the environment-oriented data are probably better monitored in terms of absolute summary values for this purpose. Lastly, it should be noted that further research is needed to improve the methods described in the Papers 1 through 5, as well as to verify their utility in multiple different herds, before any commercial implementations can be considered.

## Sammendrag


Denne afhandling bør betragtes i lyset af to primære sammenhænge, nemlig den projektmæssige sammenhæng og den samfundsmæssige sammenhæng.

I forhold til den projektmæssige sammenhæng skal det siges at det PhD projekt, der udgør grundlaget for denne afhandling, var en del af et større projekt kaldet *PigIT - improving welfare and productivity in growing pigs using advanced ICT methods*. Dette overordnede projekt dækkede flere delprojekter, der hver især beskæftigede sig med et antal forskellige fokusområder. Det overordnede mål med dette større projekt er, kort fortalt, på én gang at forbedre dyrevelfærden og produktiviteten af slagtesvin. Dette gøres ved at implementere eksisterende sensorteknologier til regelmæssigt og helst automatisk at indsamle data med henblik på monitorering. Målet med det PhD projekt, der bliver præsenteret i denne afhandling, var at kombinere det indsamlede data og, ved brug af modellering og klassifikationsmetoder, at blive i stand til at forudsige eller detektere uønskede hændelser på baggrund af disse data.

I forhold til den samfundsmæssige sammenhæng er det velkendt, at dyrevelfærd er en problemstilling som den europæiske befolkning går meget op i. Dette er særligt tilfældet i Skandinavien, inklusiv Danmark. Det er på den anden side også velkendt at den gennemsnitlige europæiske forbruger ikke er indstillet på at betale ekstra for produkter, der lover bedre dyrevelfærd. Hvis produktiviteten kan forbedres samtidig med at velfærden forbedres eller opretholdes, ville denne forbrugerholdning blive et mindre væsentligt problem. Et andet vigtigt samfundsmæssigt problem er den stigende fremkomst af antibiotikaresistente bakterier. Disse udvikler sig naturligt som reaktion på anvendelse af antibiotika, inklusiv det antibiotikaforbrug der finder sted i moderne landbrug. Det virker rimeligt at antage at en forbedring af dyrenes sundhed vil resultere i et mindsket behov for at bruge antibiotika, hvilket i sidste ende ville være med til at bremse denne evolutionære proces.

To hypoteser har været motivationen for den forskning, der vil blive diskuteret i denne afhandling, nemlig *miljøhypotesen* og *normalitetshypotesen*. Kort fortalt postulerer miljøhypotesen, at det miljø som grisene oplever, inklusiv det meget lokale miljø i fx den enkelte sti, påvirker grisenes sundhed og komfort. Ved at monitorere det lokale miljø vil man således kunne forudsige uønskede begivenheder. Normalitetshypotesen postulerer, at så længe dyrene er sunde og komfortable, vil deres adfærd og fysiologiske karakteristika kunne forudsiges af en dynamisk model, der er designet til at beskrive denne normale tilstand. Det betyder til gengæld at når dyrene er syge eller oplever ubehag, vil den samme model ikke længere være i stand til at forudsige dyrenes adfærd og fysiologi. Således vil unøjagtighederne af en sådan models kunne bruges til at advare om begyndende problemer.

De fem artikler, der bliver diskuteret i denne afhandling, tager udgangspunkt i de ovenfornævnte hypoteser. Tilsammen viser de en gradvis fremgang mod målene for det her beskrevne PhD projekt, nemlig at kombinere data fra flere kilder med det formål at forudsige eller detektere uønskede begivenheder i slagtesvin:

**Artikel 1** indikerer at monitorering af miljøet i de enkelte stier vil give nyttig information om dyrenes helbred og velfærd.  Det bør bemærkes at hverken direkte miljøobservationer af de enkelte stier eller registreringer af sundheds- og velfærdsproblemer var tilgængelige. Derfor måtte alternative markør-data bruges i stedet for.

**Artikel 2** viser at daglige opsummeringer af automatisk indsamlet temperaturdata fra de enkelte stier kan levere informationer, som er direkte anvendelige til at forudsige diarre og stivending i den enkelte sti.

**Artikel 3** viser at meget forskellige typer datastrømme kan kombineres meningsfyldt ved brug af en multivariat dynamisk lineær model (DLM). Med forskellige typer datastrømme menes data, der opsamles via forskellige sensorer, der måler forskellige variable (i Artikel 3 var der tale om vandforbrug, foderforbrug, og grisenes kropsvægt) med forskelligartede numeriske værdier, varianser, og observationsfrekvenser. Derudover præsenteres der i Artikel 3 en metode til at forene de forudsigelsesafvigelser, der forekommer for hvert observationstrin i DLM'en. Metoden til at opnå denne forening er baseret på Cholesky dekomponering. Kombinationen af en multivariat DLM til at modellere flere datastrømme og den Cholsky-baserede metode til at forene forudsigelsesafvigelserne kaldes *DLM/Cholesky* metoden.

**Artikel 4** viser at DLM/Cholesky metoden kan bruges til at forudsige uønskede hændelser i enkelte stier i et større datasæt fra en kommerciel dansk svineproducent. Metoden kan dog ikke skelne mellem forskellige typer af uønskede hændelser. Det bør i øvrigt bemærkes at kun registreringer af diarre og stivending var tilgængelige til at evaluere metodens performance. Ydermere demonstrerer Artikel 4 en simpel metode til at estimere den relative informationsværdi af de forskellige datastrømme.  Dette opnås ved systematisk at udelade eller inkludere de enkelte datastrømme i DLM'en, mens man samtidig estimerer den derved opnåede performance. Til sidst skal det nævnes de temperaturdata, der blev indsamlet i de enkelte stier, så ud til at indeholde langt mindre information når de blev monitoreret med en multivariat DLM sammenlignet med den opsummeringsmetode, der blev brugt i Artikel 2.

**Artikel 5** viste at naïv Bayesiansk klassifikation kunne bruges til at kombinere både forudsigelsesafvigelserne fra den multivariate DLM og kategoriske data der ikke blev opsamlet med sensorer, med det formål at detektere uønskede begivenheder. Denne metode blev brugt til at detektere yverbetændelse i malkekøer fra Florida.

Alle fem artikler viser, at regelmæssigt indsamlede data fra stierne eller de enkelte dyr kan omsættes til nyttig information, hvis man ønsker at detektere eller forudsige uønskede begivenheder. Tilsammen viser de desuden, at en multivariat DLM er en brugbar metode til at monitorere data, der er direkte relateret til dyrene. Derimod kan data der relaterer til det omgivende miljø sandsynligvis monitoreres bedre ved brug af absolutte opsummeringsværdier, i hvert fald til det her relevante formål. Til sidst bør det bemærkes at yderligere forskning er nødvendig for at forbedre de metoder der er beskrevet i Artiklerne 1 til 5, samt for at verificere at de kan bruges på flere forskellige gårde, før kommercielle implementeringer kan overvejes.

# List of abbreviations

| | |
|---|---|
| ADG | Average daily gain |
| ANN | Artificial neural network |
| AUC | Area under the ROC curve |
| CI | Confidence interval |
| DIM | Days in milk |
| DLM | Dynamic linear model |
| EC | Electrical conductivity (in milk) |
| EM | Expectation maximization (algorithm) |
| FCR | Feed conversion ratio |
| FN | False negative |
| FP | False positive |
| ICT | Information and communications technology |
| NBC | Naïve Bayesian Classifier |
| Neg | Negative |
| Obs | Observation |
| PLF | Precision livestock farming |
| Pos | Positve |
| RFID | Radio frequency identification |
| ROC | Receiver operating characteristics (curve) |
| SCC | Somatic cell count |
| SE | Sensitivity |
| SP | Specificity |
| TN | True negative |
| TP | True positive |
| WIM | Weeks in milk |

# Table of content

# Chapter 1: Introduction

## 1.1 Background and motivation

The PhD project described in this thesis was done as part of the larger PigIT project. The full title of this overall project is: *PigIT - improving welfare and productivity in growing pigs using advanced ICT methods*. Three words in particular stick out here: ICT, welfare, and productivity. ICT is simply an abbreviation of "information and communication technology", and productivity refers to how many pigs a farmer can produce per unit of some input factor, usually man hours. Welfare, on the other hand, is a broad umbrella term for many things, so some specification might be in order. One of the best known and most influential sets of criteria for animal welfare is the Five Freedoms, as laid out by the former advisory body to the British government known as the Farm Animal Welfare Council (2012):

1. Freedom from hunger or thirst (by ready access to fresh water and a diet to maintain full health and vigor)
2. Freedom from discomfort (by providing an appropriate environment including shelter and a comfortable resting area)
3. Freedom from pain, injury or disease (by prevention or rapid diagnosis and treatment)
4. Freedom to express (most) normal behavior (by providing sufficient space, proper facilities and company of the animal's own kind)
5. Freedom from fear and distress (by ensuring conditions and treatment which avoid mental suffering)

In the PigIT project, only items 1, 2, and 3 are actually considered so far as improving welfare is concerned, and the studies described in this thesis only relate to items 2 and 3. Specifically, the goal was to develop methods for detection and early warnings of diarrhea, pen fouling, and tail biting in growing/finishing pigs, *i.e.* slaughter pigs growing from a weight of approximately 30 kg until they reach a live weight of around 110 kg and are ready for slaughter. Tail biting was not observed in the data which were used in the papers discussed in this thesis, and so this event will not be considered further in this thesis. Pen fouling, also known as undesired excretory behavior, is the event where the pigs will start resting in the dunging area and in return excrete in the resting area. This behavior is a well known response when the temperature in the pen exceeds the limit of what the pigs will experience as comfortable (Aarnink et al. 2006), and as such this problem relates to Item 2 of the five freedoms. In addition, the act of pen fouling means increased risk of disease, thus indirectly linking it to Item 3. Lastly, pen fouling will inevitably mean costly extra work for the farm staff, as they have to clean the fouled pen, meaning that predicting and preventing pen fouling would mean a more cost efficient pig production.

Diarrhea is a common symptom of many different intestinal diseases, and the ability to achieve early warnings of this event thus relates directly to Item 3 of the five freedoms. Getting reliable early warnings about diarrhea could potentially be utilized for earlier treatment of the pigs (with antibiotics and/or other treatment options). Conventional slaughter pigs have a retention time of up to 30 days after receiving antibiotics (Videncenter for Svineproduktion 2013), during which time the farmer is not allowed to sell the pigs to be slaughtered. Thus earlier treatment of finishers would mean a quicker sale of finished pigs, thus improving the productivity. In addition, earlier detection and earlier treatment would mean less time of suffering for the afflicted animals, thus improving the welfare. Lastly, it is conceivable that that earlier detection and treatment of infectious diseases would reduce the spreading of such diseases, resulting in a long-term reduction in the use of antibiotics. In addition to being financially appealing to the farmer, a reduction of antibiotic usage is of great interest to society as a whole, as we see an ever increasing problem with antibiotic resistance in infectious bacteria (Alanis 2005; Endtz et al. 1991).

### 1.1.1 Why do we need to improve productivity?

Pig production and pig slaughter are important factors in the Danish economy, with around 20 million pigs being slaughtered in Denmark every year (The Danish Pig Research Centre 2014). Since the early 1990's, however, the relative and absolute number of Danish-born piglets being exported, mostly to Germany and Poland, has increased steadily to a point where about one third of all Danish weaned pigs (< 30 kg) are currently exported instead of being slaughtered in Denmark. As a consequence, the number of pigs being slaughtered in Denmark has been decreasing since 2003 (The Danish Pig Research Centre 2014) with slaughter houses closing and jobs disappearing as a result. During the same period, the total number of pig producers has been reduced by approximately 8 % per year, and according to projections from SEGES, Denmark will only have around 1300 pig producers by the year 2024, compared to the 3638 pig producers in business in 2015 (Videncenter for Svineproduktion 2015).

One reason for the increased export is that the foreign pig producers are able to pay more for the weaned than their Danish colleagues (The Danish Pig Research Centre 2014), reflecting the fact that raising the pigs to reach the slaughter weight can be done cheaper under *e.g.* German and Polish working conditions (Landbrug og Fødevarer 2011). In addition, the wages at *e.g.* German and Polish slaughterhouses and abattoirs are significantly lower than in Denmark, meaning that slaughtering the pigs in these countries is much cheaper than in Denmark (Landbrug og Fødevarer 2011). Therefore, in order to retain the many jobs which are dependent on production and slaughter of Danish slaughter pigs we need to make the Danish production of slaughter pigs more efficient without compromising the health and welfare of the pigs.

### 1.1.2 Why do we need to improve welfare?

While the number of slaughter pig producers has been steadily declining over the past ten years or so, the average number of pigs per producer has roughly doubled from 2969 pigs in 2003 to 5314 pigs in 2012 (Pig Research Centre 2013). Since the health of the pigs are generally assessed visually by the farm staff as they move through the herd as part of their various daily routines, it is reasonable to suspect that problems with health and comfort of the pigs can be easily missed, and that this problem only increases with increasing stock sizes. This idea seems to be supported by the findings of 2014 welfare control, conducted by the Danish Veterinary and Food Administration. They found that the most common problem for those herds, which did not pass the welfare control (27 % of all herds), was sick or injured animals, which did not receive the necessary treatment (Videncenter for Svineproduktion 2015). If we assume that these problems were generally due to oversight and not malice, they could be reduced by automated monitoring and detection systems, which could point the farm staff's attention towards those animals with particularly high risks of health issues. This assumption of oversight is in line with the result of a cross sectional study covering 20 Danish commercial pig production herds (Weber et al. 2015). The authors showed that one third of all weaned pigs, which had been assessed as healthy by the farm staff, did in fact have diarrhea when they were clinically examined.

Furthermore, an EU-wide survey conducted in 2006 showed that European citizens generally, and Scandinavians especially, consider animal welfare to be a very important issue (Eurobarometer 2007b). On a scale from 1 to 10, Danes on average rated the importance of animal welfare at 8.6., while the overall EU average was 7.8. The lowest average rating for any country was 6.9 (Lithuania and Spain). Furthermore, the high importance ratings are constant regardless of political affiliations. In addition, an overwhelming majority of Europeans (77 %) said that animal welfare either probably of defiantly needs to be improved compared to current standards. For Denmark, this number was 81 %.

Interestingly, this high level of concern for animal welfare amongst the European citizens is not reflected in their willingness to change their habits as consumers, as shown by another survey conducted mainly in 2006 (Eurobarometer 2007a). In this survey, 53 % of EU respondents said they never or very rarely consider animal welfare when buying eggs, meat, or milk, even though 74 % of these EU consumers believed that purchasing animal welfare friendly products would have a positive influence on the welfare of farm animals. Furthermore, 34 % of the EU respondents would accept no increase in price for buying welfare friendly products, while 25 % would accept up to a 5 % increase in price.

In short, it seems like the citizens in the European Union are highly concerned with the welfare of farmed animals, but that as consumers they are not willing to put in an effort or pay more to advance this cause, even though they actually believe that they have that power. It would thus seem to be the case that any improvements in animal welfare would need to be profitable for the farmer in and of itself, as consumers will not pay for it.

## 1.2 Objective and challenges of this thesis

A key element of the overall PigIT project was to utilize model based monitoring, also known as model predictive control, which is a popular strategy within the broader precision livestock farming (PLF) movement, as described by Wathes et al. (2008). The concept of model based monitoring is illustrated in Figure 1. For a physical system *e.g.* a herd, a section or pen within a herd, an individual animal, or any combination of these observational levels, data is collected using sensors, as well as in terms of diagnoses of undesired or otherwise relevant health states or behaviors. When this raw data is fed to one or more models, possibly after some pre-processing, the models should be capable of raising alarms concerning events which have already happened (detection) and/or concerning events which are likely to happen within some relevant future time frame (forewarning). These alarms can then be combined with general safety considerations and standard operating procedures for the farm to advise the farmer about what action to take.

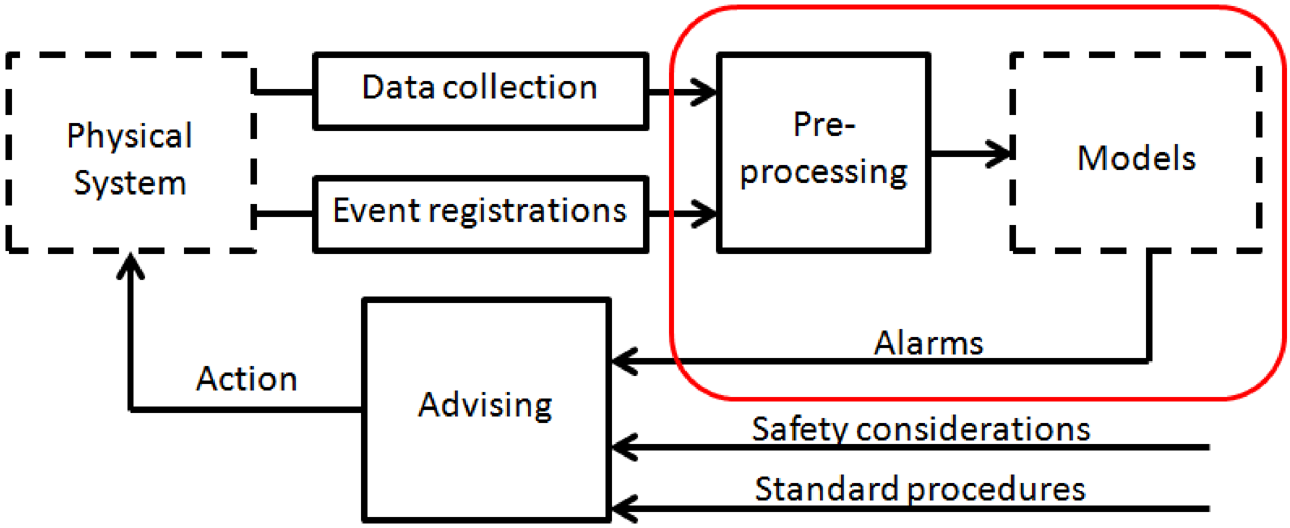



**Figure 1: The basic idea of a model-based monitoring system. The figure is adapted from Dvorak and Kuipers (1989). The red rectangle indicates the focus of the work done in relation to this thesis.**

The objective of the work done related to the papers described in this thesis was to turn the raw data into information which can be utilized for advising the farmer, as indicated by the red rectangle in Figure 1. In other words, the objective was to develop and demonstrate methods for processing collectable data into actual information pertaining to the condition of the animals. How this information should be translated into advices and decisions is however beyond the scope of this thesis, and is being covered in a parallel PhD project.

In order to meet the stated objective, several challenges had to be overcome.
One challenge is a lack of registered diagnoses of undesired events. In one of the used datasets, as described in section 2.1, no event registrations were available. In another dataset, described in section 2.2, registrations were only available for three specific kinds of undesired events, namely diarrhea, pen fouling, and tail biting. It is nevertheless known that other problems, such as influenza and pneumonia, were present in the herd, but for the research done in relation to this thesis, it was not possible to know when or where these other problems were observed. This then raised the challenge:

- How should absent or limited registrations of undesired events be handled?

Additional challenges emerged from the nature of the data being collected for the project. Overall, the data sets used in the papers relating to this thesis were collected through a multitude of sensors and could thus be very diverse; some of the collected data were numerical, while other data were categorical. Some data related to the environment experienced by the animals, some related to animal behavior, and some related to the physiology of the animals. Because of the data diversity, the values from various numerical data streams would differ in numerical magnitude and variance. Moreover, the observational frequency could differ considerably between different sensors, with some observations being made every few minutes and others being made daily or even weekly.

This diversity of data gave rise to three important challenges:

- What is the relative information value of the individual data streams within the multitude of available data streams? In other words, which data streams are most relevant to include when trying to detect and/or provide early warnings about undesired events?
- How should multiple data streams, which differ in origin, numerical magnitude, variance, and observational frequency, be combined?
- Do the various variables being measured interact with each other, and if so, how can these interactions be captured when combining the data?

For various reasons, the collected data should not be expected to be very informative in its raw form, which is why pre-processing was done. For one thing, auto-correlation in the data means that deviations from an expected pattern may be more informative than the raw values of the observations. Furthermore, data streams with high observation frequencies are likely to have a low signal/noise ratio, and might thus benefit from being aggregated to *e.g.* sums or means of longer periods. As an example, continuous water flow measurements can be aggregated to hourly sum values, thus increasing the information density of each aggregated observation. It may further be the case that deriving secondary values from a collection of primary observations, *e.g.* the rate of change within some data stream, could provide more information than merely considering the raw observations. Thus some relevant pre-processing related challenges are:

- Should the observations be condensed to a lower frequency? And if so, how?
- Are any derived values more informative than the raw observations?
- Are observed deviations from the expected pattern more informative than the raw observations?
- What is the expected pattern?

The output values, which are produced by the pre-processing, need to be passed through some model in order to determine if an alarm should be raised. In this context, the word "model" is used in a very broad sense, as it can be anything from a simple set of thresholds to more complicated classification tools. Prior to this, these models need to have been trained, *i.e.* optimized for classifying the data, based on existing observations which are known to be associated with either events or non-events.

It is important to note that some models can also be used in the pre-processing part of the model-based monitoring system, such as dynamic linear models as a method for handling the auto-correlated nature of the data streams. These pre-processing models should not be confused with the classification models discussed in this paragraph.

The alarms themselves provide additional challenges, as it should be decided if alarms should be raised before physical signs of the events are visible (forewarnings) or shortly after the events are observable (detection). Finally, the performance of these alarms must be evaluated.
All in all, the matter of classification models and their associated alarms present the following challenges:

- Which classification model should be used for raising the alarms?

- Should the system provide alarms before an event had occurred (forewarning) or afterwards (detection)?
- How should the performance of the alarms be measured?

The most important challenge of the ones listed above was probably the matter of how to best combine the various data streams. This challenge was particularly important, in part because it was the most novel aspect of the research done in relation to this thesis, but also because many of the other challenges listed above were directly tied to this challenge, and as a consequence were solved with it.

## 1.3 State of the art

In light of the research challenges listed in the previous section, an overview of the current state of monitoring pig production herds, either for the purpose of detecting diseases and other events or for the purpose of improving productivity, is in order.

The data which are usually collected in relation to herd management and precision livestock farming can be divided into two broad categories, namely animal-oriented data and environment-oriented data. In this context animal-oriented data refers to quantitative measures of the animals' behavior or their physiological traits, while environment-oriented data refer to quantitative measurements of the environment which the pigs are exposed to, such as air temperature and relative humidity. This section will first elaborate on the state of the art of utilizing these two data types for monitoring of pig production. Both common monitoring practices in modern pig production and experimental methods presented in the scientific literature will be considered. Subsequently, an overview of existing methods for combining multiple streams of data for PLF purposes will be given.

### 1.3.1 Animal-oriented data

In the practice of modern pig production, data pertaining directly to the pigs are rarely measured with sensors, automatically or otherwise. Rather, the assessment of the health and welfare of the animals is done by the farm staff through observation and interaction with the animals. For example, a commonly used sign of good health in weaned and finisher pigs is whether or not the pigs get up when the farm staff enters the section, and whether or not they generally display curiosity behavior (Eskildsen & Weber 2015). Specifically, healthy pigs of all ages should come out of hiding to examine the farm staff, while sick pigs will tend to stay hidden. In addition, the staff may observe such more or less obvious health indicators as diarrhea-like feces, which is a symptom of intestinal infections, or unclear urine, which is a symptom of bladder infection.

One noteworthy variable, which is being measured in some modern pig farms, is the eating behavior of the pigs. For example, group-housed pregnant sows are commonly fed individually using electronic sow feeders such as the Nedap Electronic Sow Feeding systems (Nedap 2013), which will recognize the individual sow using RFID transponders and feed her according to her specific feeding plan. If any sow does not eat, the system can alert the farmer, as that would be a sign of disease. Similar to this are the ACEMA systems (SKIOLD A/S 2016), which are used in breading stations such as the Danish Bøgildgård, which is run by SEGES. Like the Nedap system, the ACEMA system will recognize individual growing pigs via an RFID transponder, but will allow the pigs to feed *ad libitum*. By weighing the amount of dispensed feed and the amount left when the pig leaves the feeder, the system can monitor the exact feed consumption of the individual pig. By combining this data with regular weight measurements, the breeders can select those pigs which utilize the feed most efficiently. Lastly, some commercial slaughter pig producers use feeding systems such as those produced by Big Dutchman (Big Dutchman A/S 2016b), which will feed the growing pigs according to a predefined feeding curve, which can be adjusted during the growth period if need be. By using sensors such as the LevelCheck sensor which continuously measure the level of liquid feed in the through (Big Dutchman A/S 2016a), these systems can be made able to automatically measure how quickly the feeding through is emptied, which can provide the farmer with useful information. If, for example, the pigs are quick to finish a portion, a better growth can probably be achieved by increasing the next feed dosage, thus improving production. On the other hand, if the pigs are slow to finish the feed, it may be a sign of disease.

Several other animal-focused sensors have been tested in scientific settings, but are not yet common in commercial production. A review by Cornou and Kristensen (2013) provides a useful overview of this research. The review describes a number of animal characteristics which can be monitored using sensors, such as live weight, drinking and feeding behavior, body temperature, activity, and the sow's interest in visiting a boar.

Most of the research concerned with real-time monitoring of these characteristics focus on using the technologies to address health and welfare issues in sows. This is most notably true of various measures of sow activity which can be used to detect the onset of farrowing (Cornou and Lundbye-Christensen, 2012; Oliviero et al., 2008) as well as the onset of oestrus in sows (Freson et al. 1998). Oestrus has also been shown to be detectible via the sows interest in visiting a boar (Ostersen et al. 2010) and feeding behavior (Cornou et al. 2008), and the onset of farrowing has been shown to be detectible through changes the body temperature of the sow (Bressers et al. 1994). Feeding behavior was additionally used to detect lameness and other non-specified health-relevant conditions in group housed sows (Cornou et al. 2008).

One notable exception to the otherwise sow-focused monitoring inclination, is the use of monitoring drinking behavior in weaned pigs (4-11 weeks old) for detecting health problems with these pigs (Madsen & Kristensen 2005).

Lastly, some researchers have experimented with monitoring coughing, which can be a sign of respiratory infections regardless of the sex or age of the pigs. Specifically, some researchers have focused on developing models to distinguish between coughing sounds from healthy pigs and infected pigs (Exadaktylos et al. 2008; Ferrari et al. 2008) while others have shown that a higher frequency of coughing is positively correlated with the probability of PCR and ELISA tests for *M. hyopneumoniae* being positive (Nathues et al. 2012), although the authors advise against relying solely on the coughing index for making this diagnosis.

### 1.3.2 Environment-oriented data

Pigs are known to be particularly sensitive to their environment (Young 1981; Close et al. 2010) as well as to rapid fluctuations in that environment (Lopez et al. 1991a). It is further known that the optimal comfort temperature of the pigs will change as they grow. Specifically the optimal environmental temperature is 20$^{o}$C when the pigs weigh 15-30 kg, 18$^{o}$C when they weigh 30-60 kg, and 16$^{o}$C when they weigh above 60 kg (Kyriazakis & Whittemore 2006). If the pigs experience temperatures either above or below their comfort temperature, they will grow less efficiently; when they are cold they will spend more of the energy provided in the feed on generating body heat, and when they are hot they will have diminished appetite (Kyriazakis & Whittemore 2006), thus in both cases reducing the productivity. Furthermore, pigs are known to foul the pen in response to uncomfortably high temperatures, with the magnitude of the problem increasing with higher temperatures (Aarnink et al. 2006), and to have increased susceptibility to diarrhea-causing pathogens when the temperature is fluctuating (Shimizu et al. 1978). Moreover, if the humidity becomes too low, the pigs' mucus membrane will dry up, thereby increasing the risk of respiratory infections.

For these reasons, a detailed control of the environment within the herd is necessary in order to keep the pigs healthy and ensure a high productivity, and thus section level climate monitoring for the purpose of climate control is commonplace in modern pig production. Climate control can for example be done with integrative climate control systems such as the ones produced by the Danish company SKOV A/S (SKOV A/S 2016). Such systems will use sensors to monitor section level factors such as temperature and humidity. Readings from these sensors can then be used to automatically or manually adjust the section level environment.

For the purposes of model-based monitoring systems, it is perfectly reasonable to assume that this climate monitoring data, which is already routinely collected, could be used to infer the conditions of the pigs and thereby provide forewarnings of undesired events. By using more detailed observations, *e.g.* at the pen or the individual animal level, these warnings could be more specific. This idea is supported by the findings of Andersen et al. (2008), who showed that ear skin temperature, automatically measured for the individual pigs, was closely related to established behavioral indicators of thermal comfort.

### 1.3.3 Data combination methods in precision livestock farming

Based on the scientific literature, it seems that only little research has been aimed at combining the data and/or information produced by different sensors when it comes to detecting undesired or desired conditions in pigs. This is probably due to the fact that using animal-oriented sensors to monitor the pigs with the intention of detecting undesired events is a relatively new phenomenon, as the overview given above will attest to. A few noteworthy examples do however exist.

Cornou et al. (2008) combined sensor and non-sensor data in an attempt to achieve better detecting of oestrus, lameness, and/or other unspecified conditions for sows in a group housing system. The authors used a dynamic linear model (DLM) to model of the sows' feeding rank within the group.

The ranking of each sow was modeled separately. Information about introduction or removal of other sows was included by allowing the DLMs to adapt more rapidly in response to such changes than it otherwise would have.

Ostersen et al. (2010) combined two behavioral variables, which were being automatically monitored in sows, to detect oestrus: the frequency and the duration of visiting a boar. The duration of the visits were modeled with four distinct dynamic linear models, each describing a different scenario, including the scenario where the sow was in oestrus. By comparing the forecasted durations from each of the four models with the observed duration, combined with the prior probability of being in each of the four states, the probability of being in each of the four states could be calculated. If the oestrus model had the highest probability of being correct, an oestrus alarm was raised.

The visit frequency over periods of 6 hours was modeled using a dynamic generalized linear model. If the relative difference between observed and forecasted visit frequency was above a set threshold, an oestrus alarm was raised.

These two variables were combined using Bayes' theorem, which updates the probability of a category being true. Since a probability of oestrus was generated for each observation of the visit duration, this was used as the prior probability of oestrus. The probability of oestrus could then be updated using the presence or absence of an alarm being produced for the visit frequency model, *i.e.* the likelihood that an alarm was raised if the sow was in oestrus (sensitivity) and the likelihood that an alarm would not be raised if the sow was in oestrus (1-specificity). This combination managed to raise the sensitivity of oestrus detection sensitivity compared to what was achieved with either variable alone, while at the same time not reducing the specificity.

Many more examples of combining several different animal-oriented sensor data can be seen in relation to dairy cows, perhaps most notably in relation to detection of mastitis. A review by Hogeveen et al. (2010) gives a good overview of the modern attempts at improving the detection of dairy cow mastitis, usually by combining several lines of data using various data handling methods. Some interesting methods for handling these different data are artificial neural networks, Kalman filters, decision trees, and simple thresholds.

The most straight forward sensor combination method is the one where two sensor values are considered, and both have to surpass a set threshold in order to raise an alarm. This was the case for the threshold-based combination of electrical conductivity of milk (EC) and somatic cell count (SCC) presented by Mollenhorst et al. (2010).

More advanced data handling includes using the relevant variables as inputs in classification systems such as artificial neural networks, as done by *e.g.* Cavero et al. (2008), who used EC and milk flow as input variables. In addition to the absolute value of these variables, the authors forecasted the expected values of EC and milk flow based on the average of the last 10 milkings, a procedure commonly known as the *moving average* method. The deviations between the expected and the observed values were also used as inputs for the artificial neural network.

Even more complexity can be added by producing derived variables from primary variables. As an example, Kamphuis et al. (2010) started with five sensors, namely an EC sensor, a milk flow meter, and three color sensors (red, green, blue) for measuring the color of the milk. From these, they produced a total of 1065 derived variables, all of which were considered to be independent. From the color sensors, a simple average of the three color values served as a "combined color". From the milk flow meter, a number of milk flow-related variables could be derived, such as overall milk yield in a milking, the time from attachment of the teat cup until milk flow began, and others. From these initially derived values, even more could be derived by including various descriptor types such as mean, maximum, minimum, range, standard deviation and others, each of which was found for the sensor values during the first 500 ml, the last 500 ml, and the entire milking. All values were also compared to several different expected values calculated using moving averages of different lengths. The resulting 1065 descriptive variables were used to build a total of 16 different decision threes, which after pruning included from 5 to 55 of the descriptive variables.

In some of the above examples, the time series component of the data is addressed using the moving average method. Several alternatives to this method exist, however, for example dynamic models with Kalman filters. Kalman filters have been used by *e.g.* de Mol et al. (1997) to produce a multivariate dynamic model describing activity level, milk yield, milk temperature, and EC for individual cows, as well as a univariate dynamic model to describe the probability distribution for the amount of leftover feed per day for individual cows. These models were made for the purpose of detecting mastitis and oestrus in the individual cow. Some of the variables were considered relevant for both conditions, while others were only relevant for one condition. Just as with the moving average method, the Kalman filter would provide a forecasted value for each of the variables for each observation. These forecasts were then compared with the actual observations, and if at least one of the condition-relevant variables deviated significantly from the forecast, an alarm was raised for the relevant condition.

Lastly, a few examples exist of combining sensor and non-sensor data in order to improve mastitis detection. For example (Steeneveld et al. 2010) used a naive Bayesian network to combine sensor and non-sensor data to reduce the number of false mastitis alarms. The idea was that if a cow had been flagged as mastitis positive by an automated milking system, an updated probability of the cow actually being mastitis positive could be calculated using Bayes' theorem by considering other information, such as the parity of the cow and the season of the year.

The examples above serve simply to demonstrate that there is precedence for combining different streams of sensor data collected in livestock production for detection of relevant events, and as such the performances achieved by the cited authors are not important here. What is important is the realization that nothing in principle stands in the way of implementing any one of these methods, or any combination of them, with the intention of detection of or forewarning about undesired events in growing slaughter pigs, so long as relevant data can be collected.

## 1.4 Central hypotheses

Three central hypotheses serve as the fundamental motivations behind the work presented in this thesis. These are:

1. **The core hypothesis**, serving as motivation for the overall PigIT project:
   By systematically placing cheap sensors in the production pens of growing slaughter pigs, and by integrating the information which can be derived from these sensors, it is possible to significantly improve the production process in terms of both welfare and productivity. Thus the investment in the extra sensors will result in a net gain for the farmer.

2. **The environment hypothesis**, serving as the motivation for Papers 1 and 2:
   The local environment (*i.e.* at the pen and/or section level), to which the pigs are exposed, will affect the health and welfare of the pigs. Therefore, by monitoring the local environment, the health state and comfort of the pigs can be inferred.

3. **The normality hypothesis**, serving as the motivation for Papers 3, 4, and 5:
   When monitoring a dynamic system with a model which is optimized to describe the normal state of that system, the model will be able to accurately predict new observations, so long as the system maintain a normal state. Therefore, when the model is unable to provide accurate forecasts, the system has either changed or is in the process of changing to an abnormal state.

Please note that only the environment hypothesis and the normality hypothesis were tested in the papers described in this thesis.

## 1.5 Specific research goals

This section outlines the goals of the individual papers included in this thesis. Collectively these papers address all the challenges listed in section 1.2.

### 1.5.1 Paper 1: The effect of wind shielding and pen position on the average daily weight gain and feed conversion rate of grower/finisher pigs

In Paper 1, the overall goal was to investigate whether observing the environment experienced by finisher pigs at the pen level would be likely to provide useful information related to the health and welfare of the pigs. Since no pen level environment measurements were available, two proxies were used: whether or not the outside of the section was shielded against the wind, and the pen's distance from the entrance of the section. The implicit assumption was that the pen level environment would vary consistently with these two proxies. Furthermore, no registrations of disease or other undesired events were available. Instead the average daily gain (ADG) and feed conversion ratio (FCR) were considered the outcomes of interest.

### 1.5.2 Paper 2: Temperature as a predictor of fouling and diarrhea in slaughter pigs

With Paper 2, the intention was to evaluate the potential of pen level temperature measurements for providing forewarnings of pen level diarrhea and pen fouling. Paper 2 further served as a first attempt at condensing data streams with high observation frequency, as well for producing derived data based on primary sensor data. As described in section 2.2.1, the continuous observations were aggregated, first to 60 minute averages and then to four daily summary values for each of the two temperature sensors per pen. Lastly, Paper 2 demonstrated the use of one classification tool for raising the alarms, namely a logistic regression model.

### 1.5.3 Paper 3: A multi-dimensional dynamic linear model for monitoring slaughter pig production

In Paper 3, the goal was to demonstrate a method for combining multiple different streams of sensor data with different observational frequencies while taking the interactions between the different variables into account. It was decided to use a multivariate DLM, as described in section 3.3. Furthermore, Paper 3 presented a second classification tool for raising alarms, namely the unification of the forecast errors through Cholesky decomposition and transformation combined with a threshold value for the unified error. This combination of a multivariate DLM and the Cholesky decomposition for forecast error unification was called the *DLM/Cholesky* method.

### 1.5.4 Paper 4: A multivariate dynamic linear model for early warnings of diarrhea and pen fouling in slaughter pigs

With Paper 4, there were three main objectives: to demonstrate that the DLM/Cholesky method could be used to provide forewarnings of diarrhea and pen fouling on a large scale data set, to estimate the relative information value contained in each of the seven variables which were available at the time of the study, and to compare the performance of the DLM/Cholesky method with the summary/logistic regression method used on just temperature data in Paper 2. Paper 4 further evaluated the predictive performance of the alarms when using different prediction windows. For Paper 4, only registrations concerning diarrhea and pen fouling were available, and so only these events were considered when evaluating the performance.

### 1.5.5 Paper 5: Bayesian integration of sensor information and a dynamic linear model for prediction of dairy cow mastitis

Paper 5 introduced a third classification tool for raising alarms, namely a naïve Bayesian Classifier (NBC). This tool served as a way of combining the information from sensor and non-sensor data. The sensor data were processed via a multivariate DLM into forecast errors. These forecast errors were categorized into four categories, based on their direction (positive error or negative error) as well as their magnitude compared to the forecast variance (more or less than one standard deviation from 0). This categorization allowed an easy combination with the categorical non-sensor data by means of the NBC. This combination of a multivariate DLM and an NBC was called the *DLM/NBC* method. Paper 5 further served to demonstrate that a DLM-based method could be used to detect undesired events in animals other than pigs, as it was used to detect mastitis in dairy cows.

## 1.6 Chapter 1 references

Alanis, Alfonso J. 2005. “Resistance to Antibiotics: Are We in the Post-Antibiotic Era?” *Archives of Medical Research* 36(6):697–705.

Andersen, H. M. L., Erik Jørgensen, Lise Dybkjær, and Bente Jørgensen. 2008. “The Ear Skin Temperature as an Indicator of the Thermal Comfort of Pigs.” *Applied Animal Behaviour Science* 113(1-3):43–56.

Big Dutchman A/S. 2016a. “LevelCheck Determines the Capacitive Level in the Trough Accurately to the Last Millimetre.” Retrieved February 12, 2016 (https://www.bigdutchman.com/en/nc/pig-production/news/detail/levelcheck-determines-the-capacitive-level-in-the-trough-accurately-to-the-last-millimetre.html).

Big Dutchman A/S. 2016b. “Pig Production.” Retrieved February 10, 2016 (https://www.bigdutchman.com/en/pig-production/news.html).

Bressers, H. P. M., J. H. a te Brake, M. B. Jansen, P. J. Nijenhuis, and J. P. T. M. Noordhuizen. 1994. “Monitoring Individual Sows: Radiotelemetrically Recorded Ear Base Temperature Changes around Farrowing.” *Livestock Production Science* 37(3):353–61.

Cavero, D., K. H. Tölle, C. Henze, C. Buxadé, and J. Krieter. 2008. “Mastitis Detection in Dairy Cows by Application of Neural Networks.” *Livestock Science* 114(2-3):280–86.

Close, W. H., R. P. Heavens, and D. Brown. 2010. “The Effects of Ambient Temperature and Air Movement on Heat Loss from the Pig.” *Animal Production* 32(01):75–84.

Cornou, Cécile, and Anders Ringgaard Kristensen. 2013. “Use of Information from Monitoring and Decision Support Systems in Pig Production: Collection, Applications and Expected Benefits.” *Livestock Science* 157(2-3):552–67.

Cornou, Cécile, and Søren Lundbye-Christensen. 2012. “Modeling of Sows Diurnal Activity Pattern and Detection of Parturition Using Acceleration Measurements.” *Computers and Electronics in Agriculture* 80:97–104.

Cornou, Cécile, Jens Vinther, and Anders Ringgaard Kristensen. 2008. “Automatic Detection of Oestrus and Health Disorders Using Data from Electronic Sow Feeders.” *Livestock Science* 118(3):262–71.

Dvorak, Daniel, and Benjamin Kuipers. 1989. “Model-Based Monitorin G of Dynamic Systems.” *Proceedings IJCAI-89, Detroit, MI* 1238–43.

Endtz, H. P. et al. 1991. “Quinolone Resistance in Campylobacter Isolated from Man and Poultry Following the Introduction of Fluoroquinolones in Veterinary Medicine.” *Journal of antimicrobial chemotherapy* 27(2):199–208.

Eskildsen, Maria, and Andreas Vest Weber. 2015. *Svineproduktion*. 2nd ed. Århus N: SEGES Forlag.

Eurobarometer. 2007a. *Attitudes of Consumers towards the Welfare of Farmed Animals Wave 2*. Retrieved (http://ec.europa.eu/food/animals/docs/aw_arch_hist_sp_barometer_fa_en.pdf).

Eurobarometer. 2007b. *Attitudes of EU Citizens towards Animal Welfare*. Retrieved (http://ec.europa.eu/public_opinion/archives/ebs/ebs_270_en.pdf).

Exadaktylos, V., M. Silva, J. M. Aerts, C. J. Taylor, and D. Berckmans. 2008. “Real-Time Recognition of Sick Pig Cough Sounds.” *Computers and Electronics in Agriculture* 63(2):207–14.

Farm Animal Welfare Council. 2012. “The Five Freedoms.” Retrieved February 11, 2016 (http://webarchive.nationalarchives.gov.uk/20121007104210/http:/www.fawc.org.uk/freedoms.htm).

Ferrari, Sara, Mitchell Silva, Marcella Guarino, Jean Marie Aerts, and Daniel Berckmans. 2008. “Cough Sound Analysis to Identify Respiratory Infection in Pigs.” *Computers and Electronics in Agriculture* 64(2):318–25.

Freson, L., S. Godrie, N. Bos, J. Jourquin, and R. Geers. 1998. “Validation of an Infra-Red Sensor for Oestrus Detection of Individually Housed Sows.” *Computers and Electronics in Agriculture* 20(1):21–29.

Hogeveen, Henk, Claudia Kamphuis, Wilma Steeneveld, and Herman Mollenhorst. 2010. “Sensors and Clinical Mastitis--the Quest for the Perfect Alert.” *Sensors (Basel, Switzerland)* 10(9):7991–8009.

Kamphuis, C., H. Mollenhorst, a. Feelders, D. Pietersma, and H. Hogeveen. 2010. “Decision-Tree Induction to Detect Clinical Mastitis with Automatic Milking.” *Computers and Electronics in Agriculture* 70:60–68.

Kyriazakis, Ilias, and Colin T. Whittemore. 2006. “Environmental Management of Pigs.” Pp. 533–40 in *Whittemore’s Science and Practice of Pig Production*. Oxford: Blackwell Publishing Ltd.

Landbrug og Fødevarer. 2011. *Svinesektorens Konkurrenceevne*. Retrieved (https://www.lf.dk/lf/Tal og Analyser/Analyser/Konkurrenceevne).

Lopez, J., G. W. Jesse, B. A. Becker, and M. R. Ellersieck. 1991. "Effects of Temperature on the Performance of Finishing Swine : II . Effects of a Cold , Diurnal Temperature on Average Daily Gain , Feed Intake , and Feed Efficiency." *Journal of animal science* 69:1850–55.

Madsen, Thomas Nejsum, and Anders Ringgaard Kristensen. 2005. "A Model for Monitoring the Condition of Young Pigs by Their Drinking Behaviour." *Computers and Electronics in Agriculture* 48(2):138–54.

De Mol, R. M., G. H. Kroeze, J. M. F. H. Achten, K. Maatje, and W. Rossing. 1997. "Results of a Multivariate Approach to Automated Oestrus and Mastitis Detection." *Livestock Production Science* 48:219–27.

Mollenhorst, H., P. P. J. van der Tol, and H. Hogeveen. 2010. "Somatic Cell Count Assessment at the Quarter or Cow Milking Level." *Journal of dairy science* 93(7):3358–64.

Nathues, Heiko, Joachim Spergser, Renate Rosengarten, Lothar Kreienbrock, and Elisabeth Grosse Beilage. 2012. "Value of the Clinical Examination in Diagnosing Enzootic Pneumonia in Fattening Pigs." *Veterinary Journal* 193(2):443–47.

Nedap. 2013. "Nedap Livestock Management." Retrieved February 10, 2016 (http://en.nedap-livestockmanagement.com/).

Oliviero, Claudio et al. 2008. "Using Movement Sensors to Detect the Onset of Farrowing." *Biosystems Engineering* 100(2):281–85.

Ostersen, T., C. Cornou, and A. R. Kristensen. 2010. "Detecting Oestrus by Monitoring Sows' Visits to a Boar." *Computers and Electronics in Agriculture* 74(1):51–58.

Pig Research Centre. 2013. *Annual Report 2013*. København V. Retrieved (http://vsp.lf.dk/Om_os/Aarsberetninger VSP.aspx).

Shimizu, M., Y. Shimizu, and Y. Kodama. 1978. "Effects of Ambient Temperatures on Induction of Transmissible Gastroenteritis in Feeder Pigs." *Infection and Immunity* 21(3):747–52.

SKIOLD A/S. 2016. "ACEMA 64." Retrieved February 10, 2016 (http://acemoxsl.p5alias.domicile.fr/produit-ACEMA64-en-stalles.html).

SKOV A/S. 2016. "SKOV Climate for Growth." Retrieved February 10, 2016 (https://www.skov.com/EN/Pages/Default.aspx).

Steeneveld, W., L. C. van der Gaag, W. Ouweltjes, H. Mollenhorst, and H. Hogeveen. 2010. "Discriminating between True-Positive and False-Positive Clinical Mastitis Alerts from Automatic Milking Systems." *Journal of dairy science* 93(6):2559–68.

The Danish Pig Research Centre. 2014. *Annual Report 2014*. København V: The Danish Pig Research Centre. Retrieved (http://vsp.lf.dk/Om_os/Aarsberetninger VSP.aspx).

Videncenter for Svineproduktion. 2013. *Guidelines on Good Antibiotic Practice*. København V. Retrieved (http://vsp.lf.dk/~/media/Files/PDF - Viden/PDF - Til staldgangen - God antibiotika - DK-UK-RUS/Marts 2013 - UK/Samlet_manual-UK.ashx).

Videncenter for Svineproduktion. 2015. *Årsberetning 2015*. København V: SEGES Videncenter for Svineproduktion. Retrieved (http://vsp.lf.dk/Om_os/Aarsberetninger VSP.aspx).

Wathes, C. M., H. H. Kristensen, J. M. Aerts, and D. Berckmans. 2008. "Is Precision Livestock Farming an Engineer's Daydream or Nightmare, an Animal's Friend or Foe, and a Farmer's Panacea or Pitfall?" *Computers and Electronics in Agriculture* 64(1):2–10.

Weber, Nicolai et al. 2015. "Occurrence of Diarrhoea and Intestinal Pathogens in Non-Medicated Nursery Pigs." *Acta Veterinaria Scandinavica* 57(1):64.

Young, B. A. 1981. "Cold Stress as It Affects Animal Production." *Journal of Animal Science* (1):154–63.

Aarnink, A. J. A., J. W. Schrama, M. J. W. Heetkamp, J. Stefanowska, and T. T. T. Huynh. 2006. "Temperature and Body Weight Affect Fouling of Pig Pens." *Journal of Animal Science* 84:2224–31.

# Chapter 2: Data

For the research described in this thesis, data sets from a total of three different farms were used: Bøgildgård, a commercial Danish pig farm, and the University of Florida Dairy Herd. Here, these data sets and the herds they originate from, are described.

## 2.1 Bøgildgård (Paper 1)

Bøgildgård is a Danish boar testing facility. The data used in Paper 1 were collected in the grower/finisher station, where the pigs would grow from approximately 30 to 100 kg, before being selected either for slaughter or for use as breeding animals. The layout of the finishing station can be seen in Paper 1, Figures 1 and 2. The station housed three pure breeds of pigs: Duroc, Yorkshire and Danish Landrace, with Duroc making up approximately 50 % of the total population, and Yorkshire and Danish Landrace making up approximately 25 % each. A total of 961 groups of pigs (11-14 pigs per group) were included for Paper 1.

The pigs were fed using ACIMA-48 feeding stations, as seen on Figure 2A. One such feeding station was installed in each pen, where it would dispend dry pelleted feed *ad libitum*. Only one pig could be fed at a time, and the individual pigs were identified by the feed station via RFID ear tags. By weighing the amount of dispensed feed as well as the amount of uneaten feed left by the pig after eating, the amount of feed consumed by the individual pigs was automatically recorded.

Once per week, all pigs were manually weighed using the pig scale seen in Figure 2B. Using the RFID ear tags, the weights were recorded for the individual pigs. On the same days, the thickness of the pigs back fat was also recorded, but these data were not included in the studies presented in this thesis.

A) B)

**Figure 2: Data collection in the Bøgildgård grower/finisher section. A) The ACIMA-48 feeding station, used to record the amount of feed consumed by the individual pigs. B) The pig scale, used to manually weigh the individual pigs. The individual pigs were identified with RFID ear tags.**

Of the weighing data, only the first and last weighing of each pig were used for the purposes of Paper 1. For the feed, the sum of total feed consumed for each pig over the growing period was used. Using this information, the average daily gain (ADG) and the feed conversion ratio (FCR) over the entire growth period could be calculated for the individual pigs using equations 2.1 and 2.2, respectively. In short, FCR is a measure for the efficiency by which the pig is able to convert the consumed feed into growth.

$$ADG = \frac{Weight_{End} - Weight_{Insertion}\ (kg)}{Age_{End} - Age_{Insertion}\ (days)} \tag{2.1}$$

$$FCR = \frac{Total\ feed\ consumption\ (kg)}{Weight_{End} - Weight_{Insertion}\ (kg)} \tag{2.2}$$

The ADG and FCR were then aggregated to group average levels. These aggregations were done according to equations 2.3 and 2.4, respectively.

$$Group\_ADG = \frac{\sum_{i=1}^{i=N_{group}} ADG_i}{N_{group}} \tag{2.3}$$

$$Group\_FCR = \frac{\sum_{i=1}^{i=N_{group}} FCR_i}{N_{group}} \tag{2.4}$$

In equations 2.3 and 2.4, $ADG_i$ and $FCR_i$ are the individual ADG and the FCR values for the $i^{th}$ pig in a given group while $N_{group}$ is the total number of pigs in that group.

Missing data was not an issue in the Bøgildgård dataset.

## 2.2 A commercial Danish pig farm

For Papers 2, 3, and 4, data from a single commercial Danish pig farm were used. Specifically, the data were collected in the farm's finisher unit, housing pigs while they grew from approximately 30 kg to 110 kg, after which they were sold for slaughter. Each pen would hold 18 pigs at insertion, sorted by sex and size. For all three papers, the used data were collected between November 20$^{th}$ 2013 and December 12$^{th}$ 2014 during which three new batches were inserted into the pens.

As is seen on Figure 3, the finisher unit is made up of five sections, each containing 28 pens, corresponding to 14 double pens. Specifically the term double pen refers to two neighboring pens, which share the same feed and water supply, as illustrated in Paper 3 Figure 1A and 1B. For the PigIT project, a number of sensors were installed for data collection in four pens, *i.e.* two double pens, in four of the five sections, as highlighted of Figure 3. The data being collected were temperature, water flow, dispensed amount of feed, humidity, and live weight.

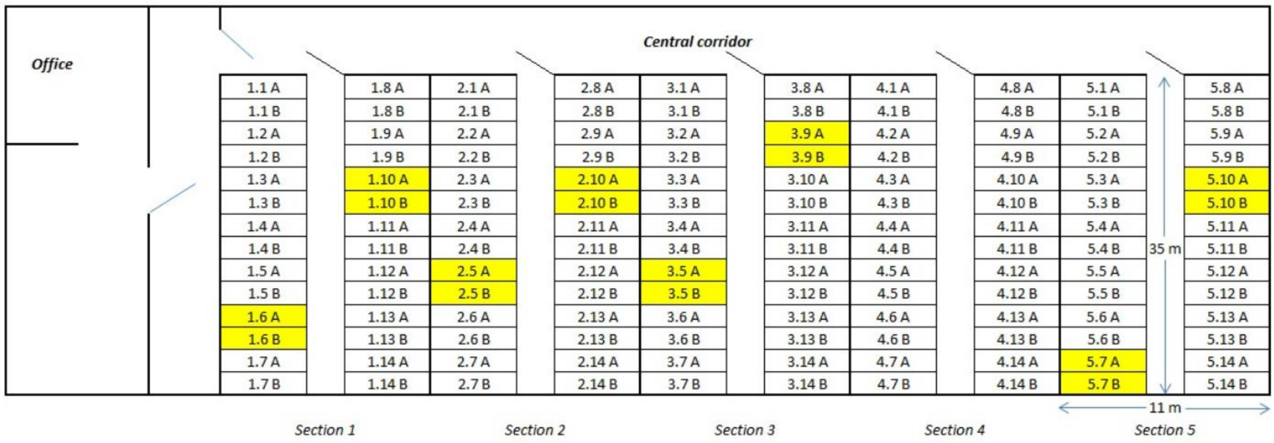



**Figure 3: Layout of the finisher unit of the commercial Danish pig farm, which provided the data for Papers 2, 3, and 4. The unit consists of five sections, each with 28 pens (14 double pens). Data were only collected in four of the five sections, and only in four pens (two double pens) per section. The pens, in which data were collected, are highlighted with yellow. Source: Krogsdahl (2014)**

The temperature data were collected on pen level. This was done using two thermometers per pen; one by the section corridor and one by the back wall of the pen, as shown in Paper 4 Figure 1A. Each thermometer would measure and record the temperature approximately once every five minutes.

Water flow was measured continuously on double pen level using flow meters installed in the water pipes, which were shared by the two pens in a double pen, as seen in Paper 3 Figure 1B and Paper 4 Figure 1C. Thus a total of eight flow meters were installed. Each of these was calibrated manually to allow for double pen-specific conversion of flow units to liters.

Liquid feed was automatically dispensed every day according to a feeding curve, which was defined for each double pen. The feeding system can be seen in Paper 4 Figure 1D. The farm staff would regularly adjust the feeding curves as necessary to correspond to the pig’s actual feed consumption.

Humidity was collected on section level by the climate computer, produced by the company SKOV A/S. The climate computer is seen in Paper 4 Figure 1B.

Live weight was only measured for the pigs in section 2. These pigs were manually weighed once per week using the scale depicted in Paper 4 Figure 1E. The pigs were identified using RFID tags, allowing the weight recordings to be made for the individual pigs.

The staff at the farm would make daily manual registrations about observations of diarrhea, pen fouling, tail biting, as well as insertions or removals to or from the pens included in the PigIT project. In practice, tail biting events were hardly ever registered and were thus ignored in the studies described in this thesis. Insertion and removal registrations were used to know the exact number of pigs in a given pen on a given day.

The data collected at the commercial farm had issues with both missing and nonsensical data. The nonsensical data were temperatures registered to be below 1 $^{o}$C, relative humidity values below 1 %

and relative humidity values above 100 %. All such values were considered as being missing. These nonsensical values were taking into account in the following summaries of missing data.

The data set covered a total of 4,228 pen-days, corresponding to 100,371 hours. Of these, data on dispensed feed amount were found to be missing for 1,222 days (29 %) and humidity data were found to be missing for 1,108 days (26 %). Data on temperature by the back wall as well as by the corridor were missing for a total of 13,005 hours (13 %). Data on drinking behavior, *i.e.* water flow and water activation frequency, were particularly plagued by missing data, with a total of 56,305 hours (56 %) missing these observations. Data from the four pens in section 2, where pigs were supposed to be weighed once per week, covered a total of 144 weeks. Within these weeks, weight is actually registered a total of 112 times, meaning that weight registrations are missing for a total of 32 weeks (22 %). This corresponds to an average of two weeks per group of pigs, and is a result of the fact that the weekly weighing of a given group stopped when the first pigs of that group were sent to be slaughtered, while those pigs which were not yet big enough stayed behind to reach the appropriate body weights.

### 2.2.1 Paper specific data descriptions

As previously mentioned, the data collected at the commercial farm were used in Papers 2, 3, and 4. This subsection elaborates on which data were used in each of these papers.

For Paper 2, only temperature data were used. Pens were considered the observational unit. These data were aggregated to hourly means, and daily summary statistics of these hourly means were calculated for each of the two thermometers. These summary statistics were maximum and minimum temperatures, greatest increase in temperature between two consecutive hours, and greatest decrease in temperature between two consecutive hours.

For Paper 3, only feed amount, water flow and live weight data were used. Furthermore, only data collected in section 2, from which live weight was recorded, were included. Double pens were considered the observational unit. The amount of feed, measured in kg, dispensed per day per double pen was normalized by the total number of pigs known to be in the double pen on the relevant day. Similarly, the water flow, measured in liters, was aggregated to daily sums and normalized by the total number of pigs in the double pen. The weekly live weights measurements of the individual pigs were aggregated to double pen means. Thus the unit of feed usage was kg/pig/day with observations every day, the unit of water flow was liters/pig/day with observations every day, and the unit of live weight was kg/pig/day with observations once every seventh day.

For Paper 4, all of the monitored variables were included. Pens were considered the observational unit.

Temperature was aggregated to hourly means for each of the two thermometers per pen. Water flow was aggregated to hourly means per total number of pigs in the double-pen. Furthermore, drinking bouts frequency, *i.e.* the number of times the water nipple was activated during each hour, was derived from the raw water flow data and normalized by the total number of pigs in the double pen. This variable was made to serve as a proxy for the activity level of the pigs. Humidity was

sometimes recorded once per hour and other times once per day. During the periods where the humidity registrations came with an hour-specific time stamp, those recordings were simply used as they were. During the periods with only daily humidity observations, which were not time stamped with a specific hour, the observations were simply assumed to be made at noon. Since humidity was only recorded at section level, all pens within the same section were assumed to experience the same humidity. Feed amount was only recorded as total amount (kg) per day, and with no hour specific time stamp, so these values were also assumed to be observed at noon.
As live weight was only recorded once per week with no hour specific time stamp, these weight measurements were likewise assumed to be made at noon on the day where the weighing took place.

Thus the unit of temperature was $^{o}$C/thermometer/hour, the unit of water flow was liters/pig/hour, the unit of drinking bouts frequency was activations/pig/hour, the unit of humidity was either percent/section/hour or percent/section/day, the unit of feed amount was kg/pig/day, and the unit of live weight was kg/pig/day.

## 2.3 University of Florida Dairy Herd

For Paper 5, data from the University of Florida Dairy Unit in Hague, Florida, were used. The herd consisted of approximately 500 Holstein cows at any given time, and the data were collected between September 2008 and March 2014. Two types of data were included in this study: sensor and non-sensor data. The event of interest was clinical mastitis, where the cows could be either positive or negative. The observational unit was the individual cow.

All sensor data were collected using sensors from the company AfiMilk®, Kibbutz Afikim, Israel. The sensor data were collected while the cows were being milked in the milking parlor seen in Figure 4A. Milking happened twice per day at 6 AM and 6 PM. The data included milk yield and electrical conductivity of the milk, both of which were measured with the milk meter shown in Figure 4B on the right. In addition, the percentagewise content of fat, protein, lactose, and blood in the milk, as well as somatic cell count (SCC) were measured by the AfiLab system, as seen in Figure 4B on the left. Lastly, the body weights of the cows were measured by AfiWeigh automated scales when the cows entered the milking parlor. One of these automated scales is seen from two different angles on Figure 4C and Figure 4D.

In summary, the various sensor variables came in three distinct sensor packages, namely the milk meter, the AfiLab and the automated scales.

The non-sensor information included parity (first or later), previous mastitis treatment (yes or no), season ("warm season" is May to August; "cold season" is September to April), and days in milk (DIM from 1 to 301). All non-sensor data were considered as categorical data. In addition, SCC was treated as categorical data, with four categories: 0-200, 200-400, 400-800, 800+ (x 1000 cells/ml). All other sensor data were treated as continuous data.

A) B) C) D)

**Figure 4: A: The milking parlor with one milk meter and one AfiLab per cow being milked. B: The milk meter (right), used to collect data on milk yield and milk conductivity, and the AfiLab (left), used to collect data on milk composition and somatic cell counts. C and D) The AfiWeigh automated scale, used for weighing the cows as they entered the milking parlor, seen from two different angles.**

Positive cases of clinical mastitis were determined by the farm staff during milking. This was done by forestripping and visual observation of flakes in the milk. The cows were checked for mastitis during both morning and evening milkings, but registrations did not come with hour-specific time stamps. All mastitis observations made during the morning milkings were registered on the same day as the observations were made. All mastitis observations made during the evening milkings, however, were not registered until the next morning, and would thus appear as if they were observed a day later than they actually were.

All sensor variables had instances of missing values. This problem was most prominent in weight and SCC data, which had 15.67 % and 6.12 % missing data, respectively. The remaining sensor variables had between 1.55 % and 3.10 % missing values. Missing values was not an issue for the non-sensor data.

## 2.4 Chapter 2 references

Krogsdahl, J., 2014. Description of Kappel farm. *PigIT report*, (5). Available at: http://pigit.ku.dk/publications/PigIT-Report5.pdf.

# Chapter 3: Modeling methodologies

For the papers included in this thesis, a number of different methods have been used for modeling the data, as well as handling the outputs produced by some of the models. The sections below serve to give an overview of the methods used in the different papers. All of the methods described below were either implemented in the statistical programming language R or made using existing R functions.

## 3.1 Linear mixed models (Paper 1)

In general, linear mixed models are models with a structure as given below:

$$y = \alpha + \beta \cdot x + \gamma \cdot z \ldots + u_i + \varepsilon \qquad 3.1$$

They describe a continuous outcome variable ($y$) given some intercept value ($\alpha$) plus some functions ($\beta$, $\gamma$, …) of a number of predictive variables which can be either numerical or categorical ($x$, $z$ …), and one or more random effects ($u_i$) when these are relevant. Finally the model contains an expression of the residual variation, *i.e.* the variation which cannot be explained by the any of the other variables ($\varepsilon$). In addition, interaction effects between two or more predictive variables can be included. If a linear mixed model has been created to describe an outcome given a number of predictive variables, an ANOVA analysis can be performed to determine the predictive significance of the individual variables in terms of *p*-values. If any variables are deemed non-significant, usually at the 5 % significance level, they can be removed from the model. By iteratively performing the ANOVA analysis and removing the least significant variable until all remaining variables are significant at the desired significance level, a final model can be achieved. This simple algorithm is known as *backwards elimination*.

In R, a linear mixed model can be defined using the function lme from the nlme library. In Paper 1, this function was used to create separate models for describing ADG and FCR for each of the three pig breeds (Duroc, Yorkshire, and Danish Landrace) given the presence or absence of wind shielding (categorical), the pen's distance from the section entrance (categorical), the season of the year (categorical), and the average start weight of the pigs in the pen, compared with the overall average for the breed (numeric). Furthermore, interactions between each of the two primary predictors (shielding and distance to the section entrance) and the comparative start weight as well as the season, were included. Backwards elimination was used to reduce the models to include variables which were significant at the 5 % level. The resulting models with estimates, describing ADG and FCR, can be seen in Paper 1 Tables 2 and 3. Paper 1 Table 2 refers to the models concerned with wind shielding, while Paper 1 Table 3 refers to the models concerned with pen distance to the entrance of the section.

## 3.2 Logistic regression (Paper 2)

In general, logistic regression models are models with a structure as given in equation 3.2.

$$logit(p) = log_e\left(\frac{p}{1-p}\right) = \alpha + \beta \cdot x + \gamma \cdot z \ldots \qquad 3.2$$

In other words, the logistic regression provides a model of the log-odds for a positive outcome of a binary variable, and the model describes how much the predictive variables ($x$, $z$, …) affect these odds, positively or negatively. In R, a logistic regression can be made using the glm function, with the family-variable set to "binomial".

In Paper 2, the binary outcome in question was the observation of either of two undesired events, namely diarrhea and pen fouling, at the pen level in a commercial Danish herd of grower/finisher pigs. Initially, the predictive variables were a total of eight summary values for the two temperature sensors in each pen. These were reduced to five significant ($p < 0.05$) or borderline significant ($p < 0.1$) variables via backwards elimination, as can be seen in Paper 2 Table 1.

The log-odds for observing an undesired event can be translated into a probability of observing the event by equation 3.3.

$$p = \frac{e^{logit(p)}}{1 + e^{logit(p)}} \qquad 3.3$$

These probabilities were provided automatically by R when using the predict function, and the alarms were raised based on these probabilities, as explained in detail in section 3.6.

## 3.3 Dynamic linear models (DLMs) (Paper 3, Paper 4, Paper 5)

In the Papers 3, 4, and 5, the data were modeled using multivariate DLMs. Paper 3 served to demonstrate the implementation of a multivariate DLM for pen level sensor data in a pig herd, while Paper 4 and Paper 5 demonstrated the application of multivariate DLM for detecting undesired events at pen level in a pig herd (Paper 4) and for individual dairy cows (Paper 5). In the context of Paper 4 and Paper 5, the basic premise was the validity of the normality hypothesis presented in section 1.4. Specifically that a model, which had been designed to accurately describe the data pertaining to the healthy state of the relevant animals, would fail to accurately forecast the observed values when the health states of the animals were compromised. This assumption is corroborated, and the principle illustrated, in Paper 5 Figure 3.

In general, a dynamic linear model is defined by an observation equation and a system equation, as shown in equations 3.4 and 3.5, respectively.

$$\boldsymbol{Y}_t = \boldsymbol{F}'_t \boldsymbol{\theta}_t + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \sim N\left(\underline{0}\,, \boldsymbol{V}\right) \qquad 3.4$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N\left(\underline{0}, \boldsymbol{W}\right) \qquad 3.5$$

The observation equation describes how a vector of observations at time = $t$ ($\boldsymbol{Y}_t$) depends on an unobservable parameter vector $\boldsymbol{\theta}_t$ as well as an observational variance ($\boldsymbol{V}$).

The system equation describes how the unobservable parameter vector evolves from time = $t - 1$ to time = $t$, under the influence of a system variance ($\boldsymbol{W}$). The values in the unobservable parameter vector is continuously estimated through Kalman filtering, as described by West & Harrison (1997).

This filtering involves forecasting the observed values based on the prior estimates of the parameter vector, the prior estimates for the co-variances of the values in the parameter vector ($\boldsymbol{C}_t$), as well as the forecast co-variance matrix ($\boldsymbol{Q}_t$). When co-modeling several variables with a multivariate DLM, the co-dependencies between those variables are captured by the various co-variance matrices ($\boldsymbol{V}$, $\boldsymbol{W}$, $\boldsymbol{C}_t$, and $\boldsymbol{Q}_t$).

The (transposed) design matrix $\left(\boldsymbol{F}_t'\right)$ and the system matrix ($\boldsymbol{G}_t$), as well as the observation and system variance-covariance matrices ($\boldsymbol{V}\ and\ \boldsymbol{W}$), are defined specifically for the models they are part of, as described in detail in the Paper 3, 4, and 5.

### 3.4 Cholesky decomposition (Paper 3, Paper 4)

In Paper 3 and 4, a Cholesky decomposition was used as a means to unify the forecast errors, which were produced by the multivariate DLMs.

In general, a Cholesky decomposition can be found for a symmetric, positive definite matrix. For a symmetric, positive matrix, $\boldsymbol{A}$, the Cholesky decomposition, $chol(\boldsymbol{A})$, is defined as a lower triangular matrix, *i.e.* a symmetric matrix where all values above the diagonal are 0, which meets the criterion stated in equation 3.6.

$$\boldsymbol{A} = chol(\boldsymbol{A}) \times chol(\boldsymbol{A})^T \qquad 3.6$$

Due to the co-variances between the forecasted variables, as represented by the forecast co-variance matrix ($\boldsymbol{Q}_t$), the forecast errors will also be co-dependent, which makes calculating a single, meaningful value for the magnitude of the several errors challenging. In Papers 3 and 4, this issue was mitigated by finding the Cholesky decomposition of $\boldsymbol{Q}_t$ for each observation in the modeled data sets. This decomposed matrix was used to transform the error vectors, resulting in a vector of mutually independent values, each of which followed a standard normal distribution. This is illustrated by Figure 5 with a hypothetical example of two modeled variables. Before the Cholesky decomposition of $\boldsymbol{Q}_t$ and subsequent transformation of the error vector, the two-dimensional probability distribution is irregular, as illustrated with the tilted ellipsis. After the transformation, we see a two-dimensional standard normal distribution, as represented by a circle with its highest density at the coordinates (0,0). In the case of three observed variables, the probability distribution of the transposed error vector will be represented by a sphere, and in cases of four or more variables, it will be represented by an *n*-dimensional hyper-sphere, where *n* is the number of variables being observed. In all cases, the vector of transposed forecast errors can be considered as a point within the *n*-dimensional sphere, and a single unified value for the set of forecast errors can be calculated as the distance from (0, ... ,0) to that point.


**Figure 5: Hypothetical 2-dimensional example of the forecast error unification via Cholesky decomposition. Before transformation, the DLM forecast errors are mutually dependent, as illustrated by the tilted ellipsis. The Cholesky decomposition of the forecast variance matrix is found, and this is used to transform the forecast error vector. The values in the transformed forecast error vector are now mutually independent and will each follow a standard normal distribution, as illustrated by the circle. The transformed forecast error vector can be thought of as a point within this circle, and the distance between (0,0) and this point is taken as the unified forecast error.**

This unified error value, $d_t^2$, can be considered as the overall difference between the expected set of observations and the actual set of observations. Thus when the DLM is designed to describe the healthy situation, the unified error can be thought of as representing the distance to the healthy situation. If this difference is above a set threshold (for a sufficient number of consecutive observations), an alarm is raised.

The unified error will follow a $\mathcal{X}^2$ distribution with $n$ degrees of freedom, where $n$ is the number of variables being observed at a given time. Specific quantile values of this distribution were used to define the control lines, which determined the magnitude of the unified errors required to raise an alarm. For Paper 3, the 0.99 quantile was used, corresponding to a 1 % probability of a given observation resulting in an alarm, assuming that the situation is normal. For Paper 4, all quantile values between 0.05 and 0.95 by steps of 0.05, as well as the 0.99 quantile, were tested.

To allow for a constant control limit in response to varying degrees of freedom, $d_t^2$ was adjusted, according to equation 3.7.

$$d_{adj.t}^2 = d_t^2 \cdot \left(\frac{\mathcal{X}^2(Quantile, n_{max})}{\mathcal{X}^2(Quantile, n)}\right), \tag{3.7}$$

where $Quantile$ is the quantile value used to determine the control line and $n_{max}$ is the maximum number of variables which can be observed at any given time.

It is worth noticing that the this error unification can only raise alarms about the system being different from what is expected, and so does not provide information about what the specific problem is.

## 3.5 Naïve Bayesian Classifier (NBC) (Paper 5)

In Paper 5, an NBC was used as a way of unifying the forecast errors produced by the multivariate DLM, as an alternative to the Cholesky-derived unification of the errors described above.

In general, an NBC makes use of Bayes' theorem, as seen in its simplest form in equation 3.8.

$$p(Pos|Obs) = \frac{p(Obs|Pos) \cdot p(Pos)}{p(Obs|Pos) \cdot p(Pos) + p(Obs|Neg) \cdot p(Neg)} \quad 3.8$$

As is seen, Bayes' theorem states that the updated, or *posterior*, probability of a condition being positive after making a given observation $\left(p(Pos|Obs)\right)$ is given by the likelihood, *i.e.* the probability of making that observation under the assumption that the condition is positive ($p(Obs|Pos)$), multiplied by the prior probability of the condition being positive ($p(Pos)$). This probability is normalized by the total probability of making the observation given either condition (positive or negative), as seen in the denominator of equation 3.5. Under the naïve assumption, multiple observations (*e.g.* the forecast errors of multiple modeled sensor variables) are considered to be independent of each other. In that case, Bayes' theorem can be expanded to the form seen in equation 3.9, when a total of *n* variables are observed.

$$p(Pos|Obs_1, \ldots, Obs_n) = \frac{\sum_{i=1}^{n}\left(p(Obs_i|Pos)\right) \cdot p(Pos)}{\sum_{i=1}^{n}\left(p(Obs_i|Pos)\right) \cdot p(Pos) + \sum_{i=1}^{n}\left(p(Obs_i|Neg)\right) \cdot p(Neg)} \quad 3.9$$

The NBC had the benefit of allowing non-sensor data, such as season of the year and previous disease registrations, to be easily combined with the information from the forecast errors. An additional benefit was that by using likelihoods related to a specific undesired event, the NBC could be used to raise specific alarms of just that specific event, as opposed to the indiscriminant alarms raised with the DLM/Cholesky method. In Paper 5, the specific event of interest was mastitis in dairy cows.

Figure 6 illustrates how the likelihoods were defined for the categorical variables, taking the season of the year as an example. Two seasons were observed in the data, namely "warm" and "cold". From the overall distribution of these seasons it is seen that 34 % of all the observations were made during the warm season and 66 % of the observations were made during the cold season, as seen in the top circle diagram of Figure 6. By dividing the learning set data into the observations where the cow did have mastitis and those where no mastitis was observed, the distributions become as seen in the two lower circle diagrams in Figure 6. The interpretation is that if a cow does not have mastitis, there is a 34 % chance of being in the warm season and a 66 % chance of being in the cold season. If, on the other hand, the cow does have mastitis, there is a 47 % chance of being in the warm season and a 53 % chance of being in the cold season. In other words, $p(Warm|Pos) = 0.47$ and $p(Warm|Neg) = 0.34$, while $p(Cold|Pos) = 0.53$ and $p(Cold|Neg) = 0.66$. When these likelihoods are known, it is possible to observe the actual season and calculate an updated probability of a cow being mastitis positive, using Equation 3.5.

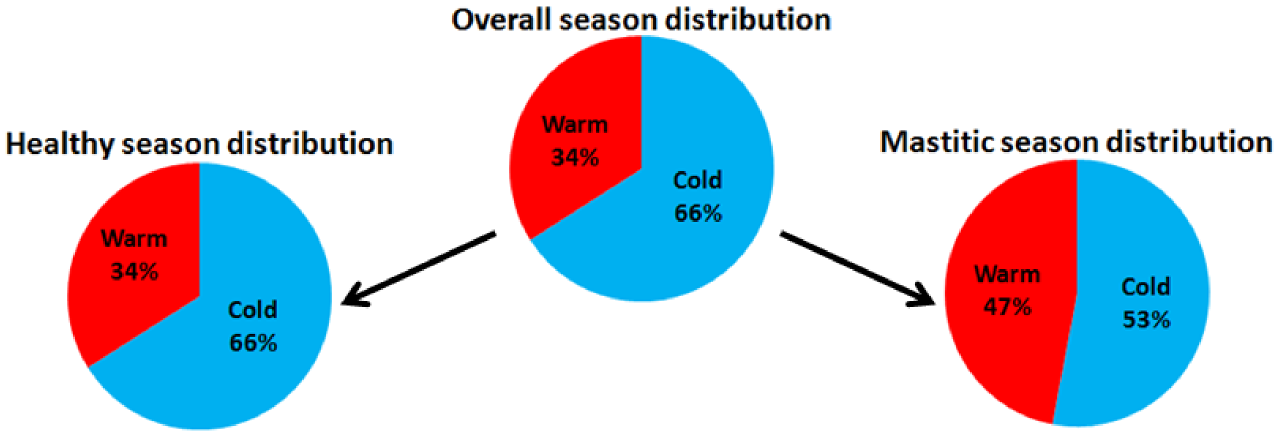



**Figure 6: Illustration of how the likelihoods of categorical observations given health states were defined, using season (warm or cold) as an example. By splitting the learning set data into observations without mastitis (healthy) and observations with mastitis, it is found that the warm season is overrepresented in the mastitis positive subset, compared to the healthy subset. Specifically, if the cow has mastitis, there is a 47 % chance of being in the warm season, whereas if the cow does not have mastitis, there is only a 34 % chance of being in the warm season.**

In the example seen in Figure 6, the healthy distribution is the same as the overall distribution. This is incidental and does not necessarily have to be the case.

For those sensor data which were modeled using a multivariate DLM, the likelihoods were defined for the forecast errors. For simplicity, each forecast error for each individual variable was assigned to one of four categories, based on how far and in what direction it deviated from zero. Figure 7 illustrates how this was done, using milk yield as an example. On Figure 7, the thick black central curve represents the forecasted milk yield on any given day of lactation between 1 and 301. All observations which are below this curve lead to negative forecast errors while all observations above this curve leads to a positive forecast error. The thinner black curves above and below the central curve represent the forecasts plus and minus one forecast standard deviation, respectively. The forecast standard deviations were calculated as the square root of the forecast variance, which is calculated as part of the Kalman filtering as mentioned above. All observations which fell more than one standard deviation below the forecasted values were categorized as "low". Observations which fell below the forecasted value but within one standard deviation were categorized as "middle low", while those which fell within one standard deviation above the forecasts were categorized as "middle high". Lastly, those observations which were more than one standard deviation above the forecasts were categorized as "high". By finding the distribution of each of these four categories in the mastitis and the non-mastitis associated subsets of the learning data, likelihoods for each of the four categories could be found, according to the same principle as illustrated in Figure 6.

It should be noted that Figure 7 shows all the milk yield observations relative to a common set of forecasts and a common set of standard deviations. This is only for illustrative purposes. In reality, each of the lactations had its own specific set of forecast values and standard deviation defined by the DLM of that specific lactation.

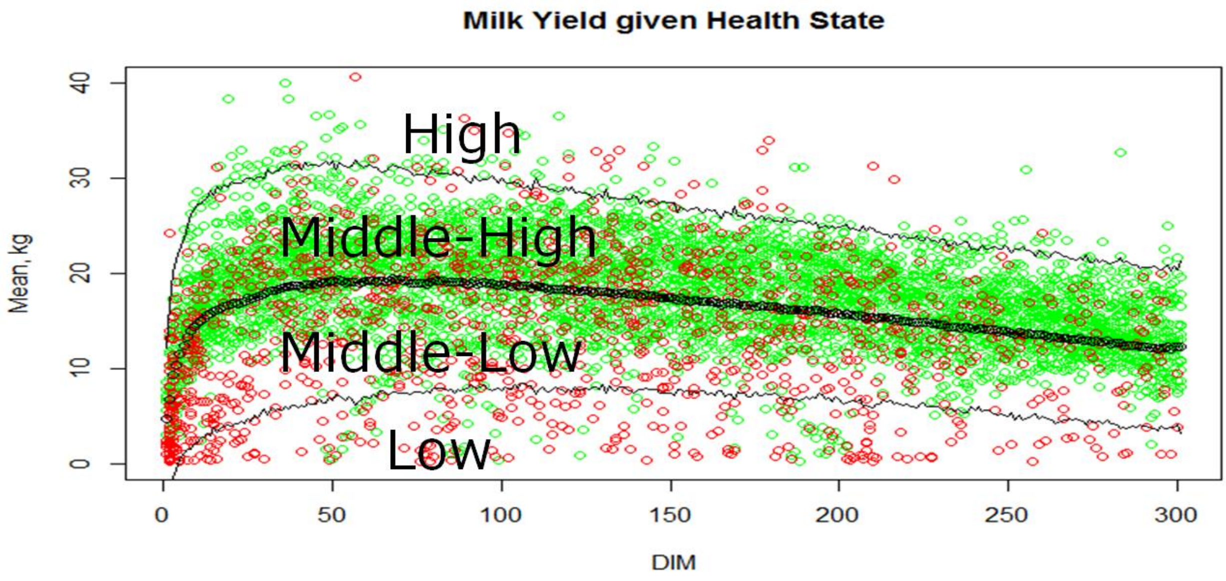



**Figure 7: Illustration of how the likelihoods were defined for those sensor variables which were modeled with the DLM. Red circles: all the mastitis-associated observations in the learning set. Green circles: all the observations not associated with mastitis in the learning set. Thick central curve: a representation of the forecasted milk yield for each day in the lactation. Thinner outer curves: representations of the forecasted values +/- one standard deviation, calculated as part of the Kalman filtering. Observations more than one standard deviation below the forecasts are categorized as "low", observations less than one standard deviation below the forecast are categorized as "middle low", observations less than one standard deviation above the forecasts were "middle high", and observations more than one standard deviation above the forecasts were "high". In reality, each of the lactations had its own set of forecasts and standard deviations, and the common values shown here are purely for illustrative purposes.**

In the example in Figure 7, it is particularly noticeable that mastitis-associated observations (red) are overrepresented in the "low" category. In fact, if the cow does have mastitis, there is a higher chance of the milk yield forecast error being "low" compared to the chance of getting this error if the cow was healthy. In Paper 5 Table 3 it is seen that the exact likelihoods of "low" forecast errors (during morning milking) are 0.10 given that the cow has mastitis and 0.02 given that the cow does not have mastitis.

## 3.6 Performance evaluation

In Paper 2, 4, and 5, the performances of the alarms were evaluated based on prediction windows around the days where undesired events were observed. The idea is illustrated in Paper 4 Figure 7 and in Paper 5 Figure 2. The prediction window stretches from a fixed number of days before the undesired event to a fixed number of days after the undesired event. If any alarms are raised within this prediction window, the window as a whole is counted as one true positive (TP). If no alarms are raised within a prediction window, the window is counted as a false negative (FN). Any days with alarms raised outside of a prediction window are counted as false positives (FP), and any days without alarms outside a prediction window are counted as a true negatives (TN). According to Hogeveen et al. (2010) this is a common method for evaluating predictive performance in the context of dairy cow mastitis, and for this reason it was decided to use it in the context of the papers described in this thesis.

For Paper 2, prediction windows of -3/+1 days were used, meaning that any alarms raised up to three days before an undesired event, or up to one day after, were counted as TP. Alarms were raised if the posterior probability of undesired events, as given by the logistic regression model, was above a set threshold.

For Paper 4, different prediction windows were tested, stretching up to six days before and up to one day after the observation of the undesired events. Furthermore, because the data used in Paper 4 was observed hourly rather than daily, any given day could in principle have up to 24 separate alarms, which in Paper 4 were referred to as *part-alarms*. Full alarms were raised if the number of consecutive part-alarms within a given day surpassed a set threshold.

For Paper 5, the data had been collected twice per day, namely at 6 AM and at 6 PM, and the morning and evening data streams had been modeled separately. For each of these two data subsets, a prediction window of -0/+0 was used. This meant that only those alarms, which were raised on the specific day where an undesired event had been observed, were counted as a TP. Alarms were raised when the posterior probability given by the NBC surpassed a set threshold.

By varying the alarm thresholds between 0 and 1 for Paper 2 and 5 (by steps of 0.001), and between 0 and 25 for Paper 4 (by steps of 1), sets of sensitivities and specificities could be found for each of the tested threshold values. Sensitivity and specificity were calculated according to equations 3.10 and 3.11, respectively.

$$Sensitivity = \frac{TP}{TP + FN} \tag{3.10}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3.11}$$

By plotting the sets of sensitivities against the sets of error rates, *i.e.* $1 - Specificity$, a receiver operating characteristics curve (ROC) (Zweig & Campbell 1993) was achieved. The area under this curve served as the primary measure of performance in all three papers.

## 3.7 Chapter 3 references

Hogeveen, H. et al., 2010. Sensors and clinical mastitis--the quest for the perfect alert. *Sensors (Basel, Switzerland)*, 10(9), pp.7991–8009.

West, M., Harrison, J., 1997. *Bayesian Forecasting and Dynamic Models* 2nd ed., New York, USA: Springer.

Zweig, M.H. & Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4), pp.561–577.

# Chapter 4: Findings

## 4.1 Paper 1: The local environment systematically affects the pigs

In Paper 1 two proxies for information about the pen level environment were evaluated as sources for information about ADG and FCR of growing pigs. These two environment proxies were whether or not outside of the section was shielded against the wind and the distance from the entrance of the section to the pen. In relation to the overall goals of the PhD project, ADG and FCR similarly served as proxies for the pigs' health and welfare.

Overall, there were no convincing effects on the FCR from neither shielding nor distance to the section entrance. Both proxies could however significantly and convincingly affect the ADG, although the effects could only be shown for Duroc pigs. In the case of shielding, this was probably due to the low number of non-Duroc groups being observed, resulting in a much lower study-power for those breeds. In the case of the pen's distance from the central corridor, this was simply because only Duroc pigs were placed at all distances, and thus the effect was only tested for these.

Shielding alone could not be shown to have a significant effect on the ADG. Significant effects were only observed when interaction effects were considered between shielding and comparative start weight ($p = 0.0002$) and between shielding and insert season ($p = 0.007$).
By comparative start weight is meant the average weight of a given group of pigs, minus the overall average start weight for the breed. On its own, the comparative start weight affects the ADG by an estimated factor of 0.011 kg, as is seen in Paper 1 Table 2. From the same table we see that not being shielded adds an extra 0.022 kg to this factor. In other words, if the goal is a high ADG, then shielding seems to be beneficial for small pigs but detrimental for larger pigs, according to the model.

By insert season is meant the season (winter, spring, summer, and autumn) at which a group of pigs is inserted into the pen. The interaction between the binary shielding variable and the four different seasons meant a total of 28 pair-wise comparisons. Of these, five comparisons included the winter/no shielding-combination being compared to some other scenario, and all of these five comparisons showed significant differences ($p < 0.05$). No other comparisons showed significant differences. In other words, the absence of shielding during winter seems to be a consistent detriment to the ADG of the pigs, while the absence of shielding at insertion during any other season did not significantly affect the ADG, all else being equal.

The distance to the section entrance, *i.e.* whether the pigs were placed in the 1$^{st}$, 2$^{nd}$, 3$^{th}$, or 4$^{th}$ pen from the entrance to the section, was found to be a significant predictor of ADG on its own, but no interaction effects with other variables could be demonstrated. Furthermore, the effect of the distance to the corridor was only significant in those pigs which were larger than their breed-average at insertion. This is likely because pigs with relatively low insertion weights generally tend to grow more slowly than larger pigs, as is seen in Paper 1 Table 1. It would thus seem like the effect of starting out small outweighs the effect of pen placement relative to the central corridor. The parameters for the model describing ADG given the distance to the corridor for larger than average pigs are seen in Paper 1 Table 3. Furthermore, Paper 1 Table 4 shows the result of a Tukey Honest Significance Difference test, which was done to determine the pair wise differences in ADG given the distance to the corridor. The 4th pen clearly yields the greatest increase in ADG relative to the 1st pen, with large pigs in the 4th pen growing 48 grams more per day ($p$ = 0.0001).

In summary, the results of Paper 1 show that the specific placements of the pigs within a herd will have significant influences on the growth of those pigs. The effects of the placement variables, in this case distance to the central corridor and the presence or absence of wind shielding, should be considered in relation to both the season of the year and the weight of the pigs at insertion. In short, larger-than-average pigs should be unshielded (except during the winter) and placed as far away from the section entrance as possible. Smaller-than-average pigs, however, should always have wind shielding, but are apparently not affected by their placement relative to the section entrance.

As previously stated, the motivating hypothesis behind Paper 1 was the environment hypothesis stated in section 1.4. This hypothesis states that the pen level environment would be able to affect the health, growth and behavior of the pigs in the same way as the overall environment has been shown to do in previously published studies, and the findings in Paper 1 are consistent with this hypothesis. It is however important to be aware of the fact that while the findings are consistent with the environment hypothesis, they do not directly show that the pen level environment is causing the described effects, since no pen level environmental observations were available.

Based on these results and considerations, it would likely be advantageous to monitor pigs at the pen level rather than at the section level, which is the current norm in the pig production industry.

## 4.2 Paper 2: Pen level temperature predicts undesired events

In Paper 2, monitoring the pen level temperature was evaluated as a method for providing forewarnings about diarrhea and pen fouling. The temperature was monitored in the form of eight daily summary values, covering two positions per pen, namely the lying area and the dunning area.

From Paper 2 Table 1 it is seen that a total of five of the included summary values were found to be significant ($p < 0.05$) or borderline significant ($p < 0.10$), and thus included in the logistic regression model, as described in section 3.2. The intercept estimate is -12.78, meaning that this is the estimated log odds for the event that either diarrhea or pen fouling is occurring in a given pen on any given day, before making any observations. This can be translated into an initial probability, $p_0$, using equation 4.1.

$$p_0 = \frac{e^{-12.78}}{1 + e^{-12.78}} = 2.8 \cdot 10^{-4}\ \% \qquad 4.1$$

It is worth noting that this value is far below the actual fraction of observed days, on which the undesired events were actually observed, which was around 1 %.

Figure 8 serves to illustrate what the estimated logit-factors in Paper 2 Table 1 actually mean in terms of the probability of observing undesired events. The figure shows how the probability of observing either of the undesired events change when the values of each of the five variables go from being at their lowest observed value to their highest observed values with steps of 0.01. From the top plot in the figure it is clearly seen that higher values of the lowest observed temperature increase the probability of observing the undesired events, bringing the probability up to nearly 100 % when the lowest temperature is above 15°C, all other things being equal. Similarly it is seen that more extreme decreases in temperature from one hour to the next increase the probability of undesired events, when those decreases are observed by the drinking nipple. In the same way, more extreme temperature increases measured by the corridor increases the risk of observing events. Interestingly, when more extreme decreases in temperature are observed by the corridor, this results in a lowered probability of observing events. This seems both counterintuitive and unlikely, given the effects we see for the other included temperature change variables, and what is known from the scientific literature, as presented in several of the papers included in this thesis. Furthermore, this variable is the only one which was only considered borderline significant with a *p*-value of 0.086, which would further legitimize a skeptical attitude towards it. Nevertheless, the model was kept as presented in Paper 2.

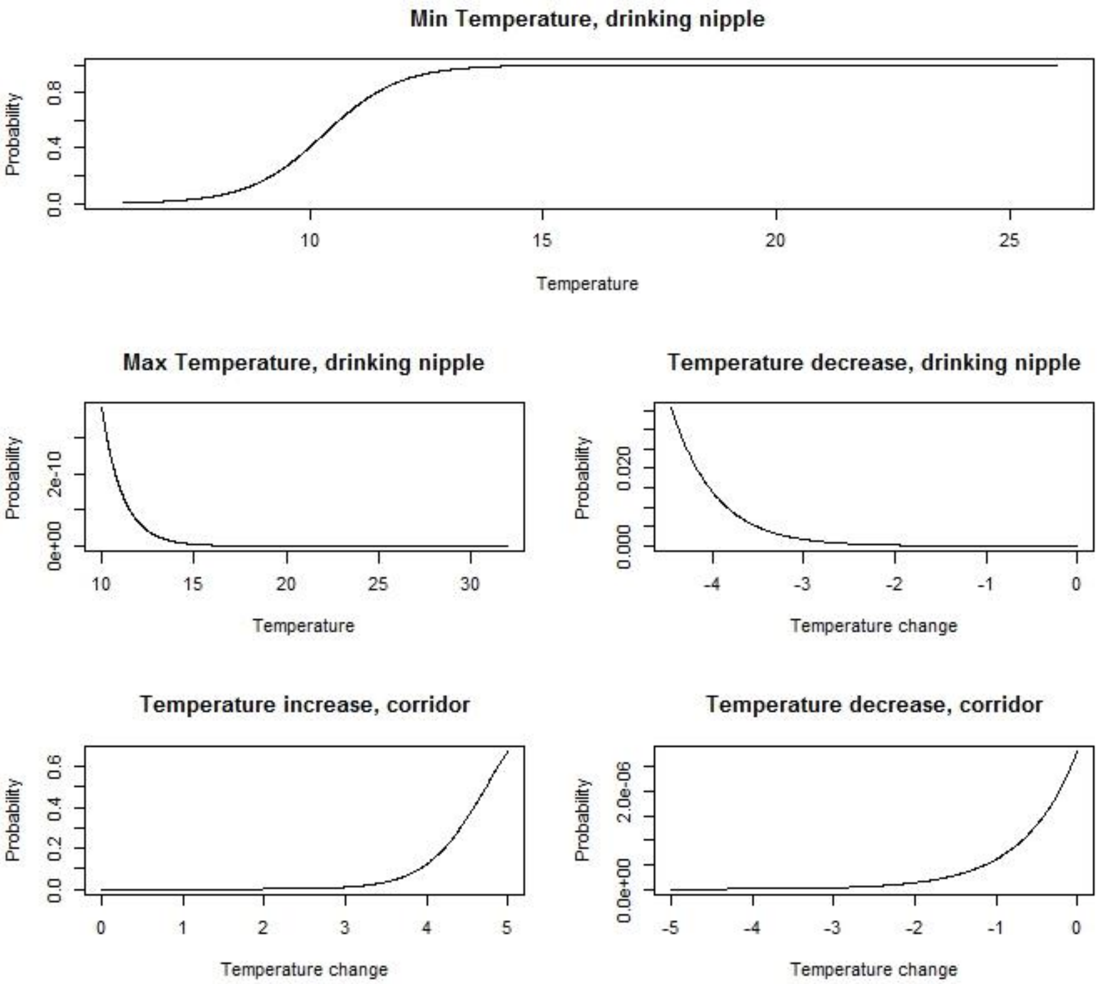



**Figure 8: Graphical representation of the effect of each of the five temperature summary variables on the probability of observing undesired events.**

When this model was tested on a test set, which did not overlap with the observations in the training set, good predictive performances were found. Figure 9 shows how the sensitivity and specificity change with the probability threshold, *i.e.* the model-given probability of an undesired event which is required before an alarm is raised. Notice that the bottom plot in Figure 9 is a zoomed in version of the top plot, made to enhance the readability. To most people it will probably seem most natural to set the probability threshold to 50 %. But, as is seen from Figure 9, this strategy would result in a sensitivity of 0 and a specificity of 1, meaning that no events will ever be detected. If one desires a specificity of 0.8, the threshold need to be set to 0.01, or 1 %. Remember that 1 % is roughly the overall prevalence of either diarrhea or pen fouling. It would thus seem that a more fruitful strategy is to set the threshold at the observed base prevalence. Similar results from Paper 5 support this notion.

The total area under the ROC curve (AUC) for indiscriminate detection of diarrhea or pen fouling was 0.80. Considering the fact that an AUC of 0.5 is expected for completely random guessing, it is clear that the pen level temperature observations actually provide real and useful information in terms of detecting these undesired events. From this, it seemed obvious that including additional pen- and higher level observations was a meaningful next step towards even higher predictive performance.

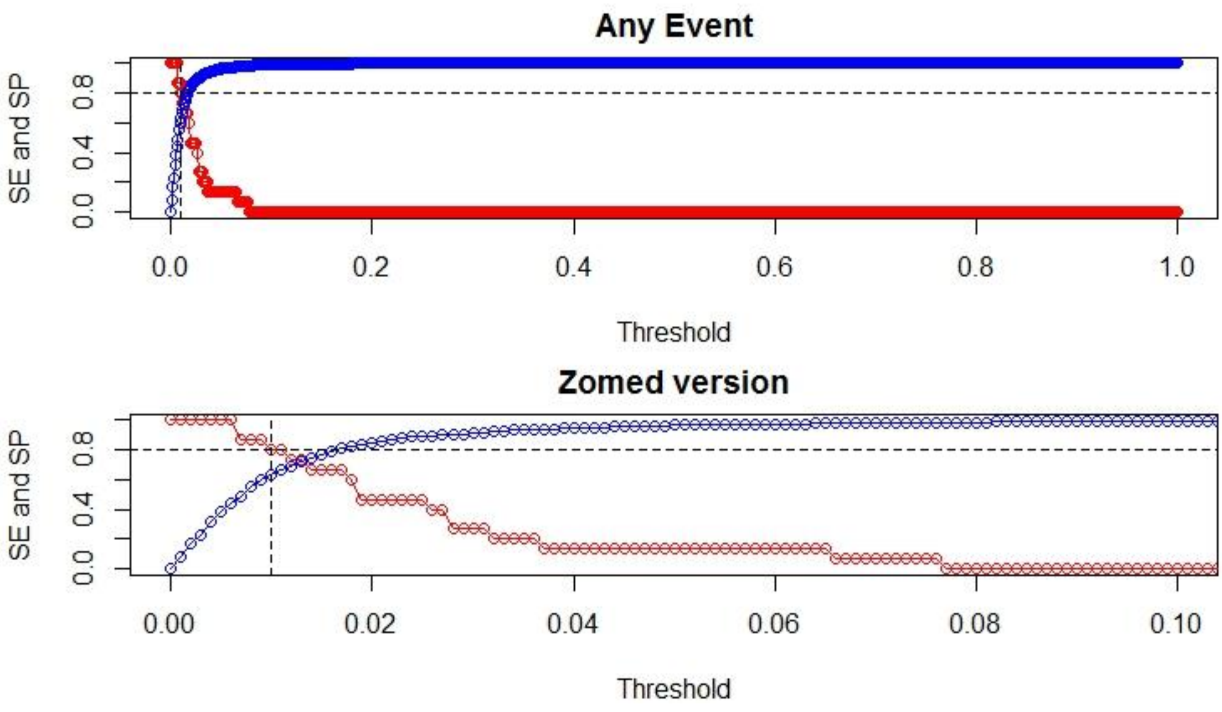



**Figure 9: The effect of probability threshold on sensitivity (SE, red) and specificity (SP, blue) of event detection. All thresholds between 0 and 1 with steps of 0.001 were tested. The bottom plot is a zomed version of the top plot, made to help the reader interpret the information more easily. An SE of 0.8 is achieved when the probability threshold is set to 0.01, or 1 %, which is also the overall prevalence of the events.**

## 4.3 Paper 3: Co-modeling diverse variables

In Paper 3, the goal was to demonstrate an implementation of a multivariate DLM for co-modeling multiple diverse variables. It was decided to co-model three observable variables, namely mean live weight of the pigs (kg), mean feed usage per pig (kg/day) and mean water flow (L/day). These three variables were chosen instead of *e.g.* pen temperature, as they are more obviously connected with each other, thus making the argument for co-modeling them easier to explain to an external audience[1]. These variables had the further distinction of being recorded with different time intervals and at different observational level.

This method was found to have a number of appealing characteristics. First of all, the multivariate DLM is able to take the co-variances between the different variables into account when forecasting the next observations. In other words, the multivariate DLM will make forecasts for each of the variables being modeled based on all available information, and not just the current value and trend of the separate variables. Secondly, the DLM makes handling missing data easy, as it will only update the values in the parameter vector, from which the next observations are forecasted, given the variables which are actually being observed at a given time. Thus the forecasts are always made given the best available information.

Paper 3 Figure 2 shows three examples of Cholesky-unified forecast errors being produced by the multivariate DLM. Two main things can be learned from this figure: first, this method is capable of producing very clear spikes in the unified error in relation to an undesired event, namely pen fouling, as seen in the middle plot. Second, the data which are relevant in relation to a given undesired event needs to be observed in order for the model to be able to produce an alarm about that event. This is exemplified in both the middle and the bottom plot. In both cases, diarrhea is observed during a longer period where data on water flow are missing, resulting in both diarrhea events going undetected.

Lastly, it is worth noting that Paper 3 marks the first time where the normality hypothesis was explicitly stated: if a model is specifically designed to accurately forecast the observations of a healthy system (be that a single animal, a whole herd, or anything in between), one would expect to see large forecast errors when the system deviates from this state. It then follows that if such a model starts producing very large forecast errors, then the system has shifted from a healthy to a non-healthy state. Thus a model optimized for describing a healthy system can be used as part of an alarm system for detecting undesired events, such as disease outbreaks.

Having demonstrated that the multivariate DLM method could be meaningfully used to co-model (double) pen level observations in a meaningful way, it made sense to test this method, and the normality hypothesis, on grander scales.

[1] Paper 3 was made as a conference paper and was presented at the EC-PLF conference in Milan, September 2015.

## 4.4 Paper 4: Evaluating the DLM/Cholesky method

In Paper 4, the goals were to test the utility of the DLM/Cholesky method for providing forewarnings of diarrhea and pen fouling on a large data set, to estimate the relative information value of all available data streams for this purpose, and to compare the performance with that achieved in Paper 2.

With respect to the first goal, an AUC of up to 0.88 was achieved, as well as a specificity of 0.81 when the sensitivity was held at 0.80. This was done using a prediction window of -5/+1 days around the days with event observations while using the 0.70 quantile of the $\mathcal{X}^2$ distribution with between 1 and 7 degrees of freedom, depending on how many variables were observed at any given time. This quantile value determined the control line for part-alarms, as described in section 3.4. Figure 10 shows the sensitivity and specificity given the threshold for number of consecutive part-alarms needed for one full alarm, as well as the corresponding ROC curve. As is seen, a sensitivity of 0.80 is achieved with a threshold of either five or six part-alarms required per one full alarm, but that a higher specificity is achieved with a threshold of six part-alarms.

The ROC curves resulting from using the same control line but other prediction windows can be seen in Paper 4 Table 9 and in the top plot of Paper 4 Figure 10.

Regarding the estimation of the relative value of the various variables, this was done by omitting one or more variables from the model to see how much this would affect the performance. The effects on the performance are illustrated in the middle plot of Paper 4 Figure 10. With the -5/+1 prediction window, data on live weight, feed usage, and humidity could be completely ignored while still maintaining an AUC of 0.88, so long as water flow and drinking bouts frequency data were still included. When these drinking behavior data were omitted, however, the inclusion of live weight, feed usage, and humidity could actually be seen to contain at least some information.

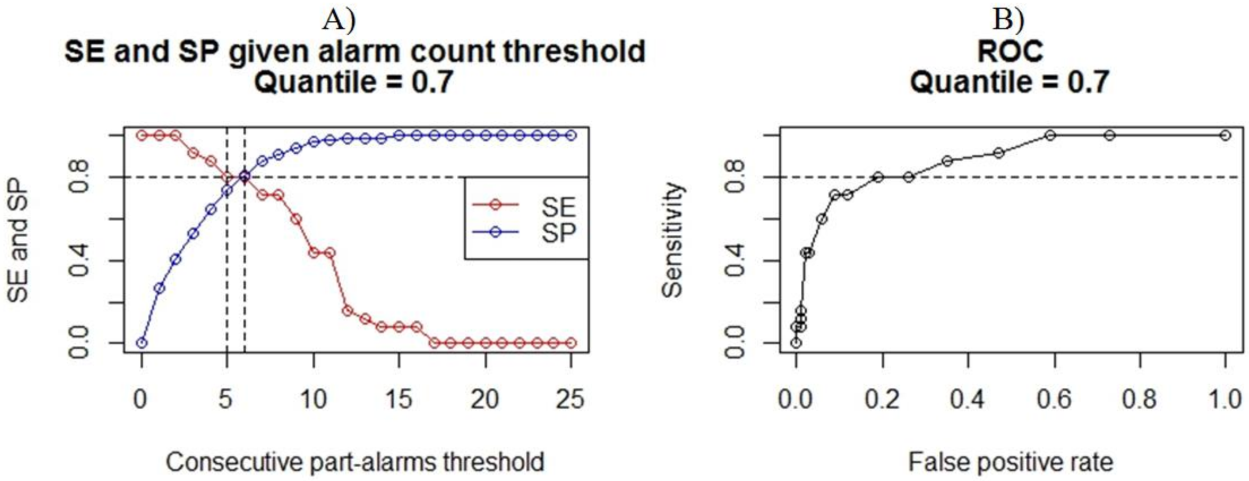



**Figure 10: A) Sensitivity (SE) and specificity (SP) given the number of consecutive part-alarms required for one full alarm. The dashed lines indicate that a sensitivity of 0.8 is achieved when the threshold is set to either 5 or 6. B): The ROC curve corresponding to the SE and SP values seen on plot A).**

Specifically, when drinking behavior was omitted, the combination of live weight, feed usage, humidity, and temperatures produced an AUC of 0.73, compared to the AUC of 0.65 achieved with the temperatures alone. This tells us that live weight, feed usage, and humidity carry very little, although some, useful information with respect to warning about diarrhea and pen fouling. This small amount of information was contained entirely in feed usage and humidity. This is at least the case given the way these data are currently collected and process. In the discussion of Paper 4, the argument is made that these variables might be more useful if they were collected more often, such as daily for the live weight, or hourly for the humidity.

Omitting either water flow or drinking bouts frequency caused moderate reduction in the performance, reducing it to 0.85 and 0.87, respectively. A much greater reduction was seen when both of these were removed at the same time, as this resulted in an AUC of 0.73. This reveals that while one of these water variables may substitute partially for the other, the complete omission of drinking behavior is severely detrimental to this models performance.

The omission of temperature measured in the lying area, but not by the corridor, also resulted in noticeable reduction of the performance, but not nearly to the same extend that omitting drinking behavior did.

All in all, a list ranking the various data streams from most to least information value, given the goal and results of Paper 4, is as follows:

1. Drinking behavior
2. Temperature (in the lying area)
3. Feed usage
4. Humidity
5. Live weight

For comparison with the model presented in Paper 2, remember that the summary/logistic regression method reached an AUC of 0.80 for indiscriminate detection of diarrhea and pen fouling with a prediction window of -3/+1. It is interesting to note that this was not the case for the purely temperature-based model made with the DLM/Cholesky method. As mentioned earlier, and as is seen in Paper 4 Table 11, this method, using the same prediction window as in Paper 2, only managed an AUC of 0.65. From Paper 2 it is of course known that much more information is contained in the temperature data than this rather disappointing performance would otherwise suggest. The natural conclusion is therefore that the DLM/Cholesky method is simply not the best way to monitor the temperature if the goal is forewarnings of diarrhea and pen fouling.

## 4.5 Paper 5: Evaluating the DLM/NBC method

The goals of Paper 5 was to demonstrate that the DLM/NBC method could be used for detecting clinical mastitis in dairy cows, and to estimate the relative values of each of the sensor packages described in section 2.3 (including the collection of non-sensor data) for this purpose.

The fact that the three ROC curves seen in Paper 5 Figure 5 have respective AUC values of 0.89, 0.85, and 0.73 is proof that the DLM/NBC method can indeed be used for detecting clinical mastitis in dairy cows.

Much like in Paper 4, the relative value of the different data packages were evaluated by testing the predictive performances achieved with the different possible combinations of the three sensor packages. In Paper 5, the predictive performances were evaluated using three different measurements: the AUC of the ROC curve (as was the case in Papers 4 and 2), the specificity achieved when sensitivity was held at 0.80, and the error rate achieved when sensitivity was held at 0.80. The results of these tests are seen in Paper 5 Table 4. One of the first things to notice is that the non-sensor data, designated as package 0, does contain some relevant information, although the effect of omitting these data are somewhat modest, but consistent over all three performance measures (change in AUC: from 0.89 to 0.88, change in specificity: from 0.81 to 0.79, change in error rate: from 0.19 to 0.21).

The next thing to notice is that the automated scale, designated as package 3, does not offer any additional information so long as both the milk meter and the AfiLab (packages 1 and 2, respectively) are included. It does however seem to add some modest information when combined alone with either the milk meter or the AfiLab. From Paper 5 Table 3 it is seen that observing a forecast error for the live weight in the "Low" category (*i.e.* the live weight is more than one standard deviation below what was expected) is more likely if the cow has mastitis than if it does not. On the other hand, this relationship is not reflected in the descriptive statistics seen in Paper 5 Table 1. Here it seems that, if anything, the mastitis positive cows tend to weigh slightly more than the cows without mastitis, although the two are well within one standard deviation from each other. It therefore seems more likely that the apparent effect of including live weight is an artifact of data noise and not in fact a true effect.

Regarding the milk meter and the AfiLab, however, the effects of omitting either or both of these packages seem undeniable. Determining which of the two is more important is a more difficult matter. Including the milk meter while omitting the AfiLab yields a slightly better AUC than the other way around (AUC = 0.86 vs. AUC = 0.85). If one wish to keep the sensitivity at 0.80, however, including only the AfiLab instead of the milk meter results in a slightly better specificity (0.74 vs. 0.73) and error rate (0.25 vs. 0.27). But by the same token as with the apparent effect of live weight, these differences are so small that they could conceivably be the result of mere noise. Thus, if a farmer was forced to buy just one of these sensor packages, the best recommendation would be to go for the cheaper option rather than base the decision on these numbers. And at any rate, the best performance would be achieved by combining both packages 1 and 2 while also including the non-sensor information.

# Chapter 5: Paper 1

Published in *Livestock science*:

Jensen, D.B., Toft, N. & Cornou, C., 2014. The effect of wind shielding and pen position on the average daily weight gain and feed conversion rate of grower/finisher pigs. *Livestock Science*, (167), pp.1–9.

Livestock Science 167 (2014) 353–361

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

# The effect of wind shielding and pen position on the average daily weight gain and feed conversion rate of grower/finisher pigs

Dan B. Jensen [a,*], Nils Toft [b], Cécile Cornou [a]

[a] HERD – Centre for Herd-oriented Education, Research and Development, Department of Large Animal Sciences, University of Copenhagen, Denmark
[b] Veterinary Institute, The Technical University of Denmark, Denmark

ARTICLE INFO

*Article history:*
Received 13 March 2014
Received in revised form
22 June 2014
Accepted 30 June 2014

*Keywords:*
ADG
Boars
FCR
Pig production
Weight gain
Wind shielding

ABSTRACT

Pigs are known to be particularly sensitive to heat and cold. If the temperature becomes too low, the pigs will grow less efficiently and be more susceptible to diseases such as pneumonia. If the temperature is too high, the pigs will tend to foul the pen, leading to additional risks of infection. Furthermore, unpublished data show that the temperature within a single section of grower/finisher pigs can vary considerably from pen to pen, and previous studies have shown that pigs can be significantly affected by wind, even when not directly exposed to it. To address this latter concern, some pig producers and research stations have implemented a shielding to prevent winds from blowing between separate sections of the pig housing buildings. However, according to our search of the literature, no published studies have ever investigated the effectiveness of such shielding.

To determine the significance of the effects of wind shielding, linear mixed models were fitted to describe the average daily weight gain and feed conversion rate of 1271 groups (14 individuals per group) of purebred Duroc, Yorkshire and Danish Landrace boars, as a function of shielding (yes/no), insert season (winter, spring, summer, autumn), start weight and interaction effects between shielding and start weight and shielding and insert season. Such a model was fitted separately to the data collected for each breed. Shielding was found to have significant interaction effects with season ($p=0.007$) and start weight ($p=0.0002$) for Duroc pigs, but no effect could be shown for Yorkshire or Danish landrace.

To determine the effect of a group's placement relative to the central corridor of a grower/finisher station, a similar model was fitted to the data for Duroc pigs, replacing shielding with distance from the corridor (1st, 2nd, 3rd or 4th pen). The effect could not be tested for Yorkshire and Danish Landrace due to lack of data on these breeds. For groups of pigs above the average start weight, a clear tendency of higher growth rates at greater distances from the central corridor was observed, with the most significant differences being between groups placed in the 1st and 4th pen ($p=0.0001$). A similar effect was not seen on smaller pigs. Pen placement appears to have no effect on feed conversion rate.

* Correspondence to: HERD – Centre for Herd-oriented Education, Research and Development, Department of Large Animal Sciences, University of Copenhagen, Grønnegårdsvej 2, 1870 Frederiksberg, Denmark. Tel.: +45 353 33021.
*E-mail address:* daj@sund.ku.dk (D.B. Jensen).

http://dx.doi.org/10.1016/j.livsci.2014.06.028

No interaction effects between shielding and distance to the corridor could be demonstrated. Furthermore, in models including both factors, the effect of distance to corridor completely dominated over the effect of shielding, suggesting that shielding should at most be considered of secondary importance.



## 1. Introduction

Pigs are known to be more sensitive to heat and cold than, for example, cattle and will thus be more likely to expend energy to maintain a constant body temperature in response to varying surrounding temperatures (Young, 1981). This means that when exposed to temperatures below a critical threshold, the animal will expend energy on maintaining a constant body temperature rather than on growth. When wind is blowing, this threshold is significantly reduced, even if the pigs are not directly exposed to the blowing wind (Close et al., 1981; Mount, 1966). This is in line with the findings of Fitzgerald et al. (2009), who showed both the outside temperature and the wind speed to be significant factors relating to the risk of pigs arriving dead or exhausted after a transport to the abattoir, despite being shielded by the walls of the truck. Fitzgerald et al. (2009) found that low temperature generally increased the risk of pigs dying, while high wind speeds increased the risk during winter and reduced the risk during summer.

In addition to metabolism and stress, the temperature to which a pig is exposed can also affect the risk of infections. For example, lower temperatures increase the risk of pigs being infected with *Mycoplasma hyopneumoniae*, the bacterium which is the main causative agent of enzootic pneumonia and porcine respiratory disease complex (Segalés et al., 2012). Aside from being a welfare problem, such infections can significantly reduce the daily weight gain of the pigs, as demonstrated by Wilson et al. (2012). Also, it is known that pigs which experience temperatures above the limits of their comfort (typically 20–25 °C) will tend to foul the pen (Aarnink et al., 2006). The pigs will thus excrete in the resting area and rest in the dedicated excretion area of the pen, which can cause a series of potential health problems. Furthermore, currently unpublished data, given to us by the Danish research centre Foulum (AU Foulum 2013), show that within the same section of a pig production building, the temperature can vary considerably between individual pens.

Given the sum of the above information, it is meaningful to assume that certain areas of a given pig production building would pose problems for the growth and health of the pigs, which we expect would be evident from a slower average growth rate and/or less efficient utilization of the feed given to the pigs in those troublesome areas. Such troublesome areas might be pens with back walls exposed to the wind and temperature of the outside environment, or pens placed near the central corridor of a production building, as these will experience more draft. Based on the first of these assumptions, the Danish Pig Research Center (Danish Agriculture and Food Council, 2013) has put up wooden walls intended to provide shielding, preventing the wind from blowing between separate sections of Bøgildgård, their fish bone-shaped research and boar breeding station in Kjellerup, Denmark (see Fig. 1A). However, from our search of the scientific literature, there seem to be no studies examining the utility of such shielding, nor of the effect on the pigs being placed near the central corridor. A better understanding of the true effect of these factors would help farmers to make informed decisions about the best placement of pigs in terms of efficient production. Thus, the purpose of this study is to assess the effect of wind shielding and pen placement relative to a station's central corridor on average daily weight gain (ADG) and feed conversion rate (FCR). These primary predictive factors are examined in combination with the effects of varying seasons and pig start weight.

## 2. Materials and methods

This study was concerned with the effect on the main outcomes (ADG and FCR) at the level of the pen. Thus, ADG and FCR were first calculated for the individual pigs in the dataset, and subsequently the data was aggregated to the level of groups. We use the term “group” to mean up to 14 individual pigs, which are sharing the same pen during the same period of time. ADG and FCR were calculated at the individual pig level according to the following equations:

$$ADG = \frac{Weight_{End} - Weight_{Insertion}\ (kg)}{Age_{End} - Age_{Insertion}\ (days)} \quad (1)$$

$$FCR = \frac{Total\ \ feed\ \ consumption\ (kg)}{Weight_{End} - Weight_{Insertion}\ \ (kg)} \quad (2)$$

The aggregation of ADG and FCR to group level was then done according to the following equations:

$$Group\_ADG = \frac{\sum_{i=1}^{i=N_{group}} ADG_i}{N_{group}} \quad (3)$$

$$Group\_FCR = \frac{\sum_{i=1}^{i=N_{group}} FCR_i}{N_{group}} \quad (4)$$

where $ADG_i$ and $FCR_i$ are the ADG and FCR values for the $i$th pig in the group respectively and $N_{group}$ is the number of pigs in the group.

### 2.1. Source and study population

This study was done using data collected at the Danish research station Bøgildgård (Danish Agriculture and Food Council, 2013), primarily used for boar breeding. The study population was purebred grower/finisher pigs (30–100 kg) in a grower/finisher station with mechanical and natural ventilations. The mechanical ventilation was placed in the ceiling of the sections and automatically controlled according to a number of set points for temperature and humidity. The windows were as a rule never open. The boars were given dry pelleted feed *ad libitum*. The feed consisted of 1.05 FUgp (Feed Units for growing pigs),[1] 16.1% crude protein, and 0.9% lysine per kilogram feed. The feed intake for each animal was individually recorded using "ACEMA-48" electronic feeding stations. The data set contains manual weight measurements and automatic feed consumption registrations from 18,525 individual pigs of three different breeds (Duroc, Yorkshire and Danish Landrace) from September 2008 to December 2011. The 18,525 individual pigs made up a total of 1473 groups. Groups of 11–14 individuals made up 93% of the observations. For this reason, only groups with a size of 11 or above were included in the study. To get realistic estimates of the average daily weight gain over a group's time in the grower/finisher station, only groups with a realistic average start weight of 40 kg or below were included in the study. Table 1 provides descriptive statistics of the study population.

The structure of the grower/finisher station at Bøgildgård is illustrated in Fig. 1. As shown, the station is laid out as a symmetrical fishbone structure with a central corridor and eight sections (sub-buildings) on each side (Fig. 1A). Each section holds four pens along each long wall, *i.e.* eight in total. For the purpose of this study, the layout of the station was considered as an ($X$,$Y$)-coordinate system with the west–east direction representing the $X$-axis and the north–south direction representing the $Y$-axis. The centre of the coordinate system is set at the centre of the central corridor. Pens east of the central corridor (vertical shaded area in Fig. 1B) are represented by positive $X$-values while pens west of the corridor are represented by negative $X$-values. Similarly, pens north of the central section, which does not house any pigs (horizontal shaded area in Fig. 1B), are represented by positive $Y$-values, while pens south of central section are represented by negative $Y$-values. The circles (Duroc), triangles (Yorkshire) and crosses (Danish Landrace) indicate which breeds are most commonly placed in the various pens. As is seen, the breeds are not placed randomly throughout the station, but are systematically put in specific pens. Neighboring sections on the same side of the corridor have open space between them. To prevent wind from blowing between the neighboring sections, wooden walls are placed between neighboring sections, as seen in Fig. 2. In Fig. 1B, this shielding is marked by thicker lines.

The Duroc groups are found most evenly distributed throughout the station, as they are generally placed in every pen on the right-hand side of the section, as seen when entering from the central corridor. Groups of Yorkshire and Danish Landrace are generally placed at the left hand side of the section with Danish Landrace groups generally being placed in the pens closer to the central corridor and the Yorkshire groups being placed at the far end of the section. Beyond the eight pens of each section is a separate room (not depicted in Fig. 1B) which is used for keeping the pigs for weighing purposes. This separate room means that the walls of the farthest pens, which run parallel with the central corridor, are not directly in contact with the outside environment.

[1] FUgp approximates 12.5–12.8 MJ metabolizable energy.

### 2.2. Data analysis

All data analysis, models and data representations were made using the free statistical software tool R, version 2.15.3 (The R Core Team, 2013).

The main outcomes of this study are ADG and FCR. The predictors of the two main outcomes are

a. Wind shielding (yes/no) and distance from the central corridor (1st, 2nd, 3rd or 4th pen) were the primary predictors of interest.
b. Start weight, *i.e.* the average weight in kg of the pigs in a given group, measured two days after insertion.
c. Insert season, *i.e.* the season (winter, spring, summer, autumn), during which the pigs were placed in the grower/finisher station which was included to account for the influence of the season at the time when the pigs are supposedly most vulnerable. The seasons were defined as follows: winter=December, January and February; spring=March, April, and May; summer=June, July, and August; and autumn=September, October, and November.

#### 2.2.1. Linear mixed models

Linear mixed models were made using the R function lme. Using ANOVA tests, the models were used to assess the significance of the predictors for ADG and FCR. Different models were fitted with respect to the two primary outcomes. Backwards elimination was used to arrive at a final model, consisting only of significant factors. Normality was checked by observing normal quantile–quantile plots of the residuals of the reduced model, using the R function qqnorm. The assumptions of normality were considered to be met if the plot presented a straight line.

*2.2.1.1. Effect of shielding.* From Fig. 1B, it is seen that the sections facing the north and south ends of the station are not shielded against the wind, as indicated by the lack of thick solid lines at these positions. To assess the effect of the shielding, the groups occupying the pens placed at these unshielded walls were compared with the groups in the shielded sections. Those groups, which were placed in the sections with unshielded walls, but not directly at the unshielded walls ($Y$-coordinates $(\pm)$ 7), were not considered in this context. Thus, in Fig. 1B, the unshielded pigs are those placed at the $Y$-coordinates $(\pm)$8, while the shielded are placed at the $Y$-coordinates $(\pm)$ 1, 2, 3, 4, 5 and 6. The unshielded subset includes 76 Duroc-, 43 Yorkshire- and 43 Danish Landrace-groups while the shielded subset

**Table 1**
Descriptive statistics of the source population. In the study population, only groups of pigs of 11 or more with an average start weight at or below 40 kg are included.

| Variable | *N* | Outcome varibles | | | | Variables included in the calculation of outcome variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ADG | | FCR | | Days in station | | Insertion weight, kg | | End weight, kg | | Feed consumption, kg | |
| | | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D | Mean | S.D | Mean | S.D. |
| Total | 1271 | 0.97 | 0.12 | 2.21 | 0.38 | 73.86 | 13.90 | 27.90 | 2.25 | 100.18 | 14.24 | 160.96 | 36.32 |
| Breed | | | | | | | | | | | | | |
| Duroc | 560 | 1.01 | 0.11 | 2.08 | 0.29 | 68.91 | 12.02 | 28.96 | 2.06 | 99.15 | 13.78 | 147.57 | 33.07 |
| Yorkshire | 353 | 0.90 | 0.11 | 2.33 | 0.46 | 79.25 | 14.82 | 26.74 | 2.07 | 99.39 | 14.95 | 169.64 | 35.48 |
| Danish Land Race | 358 | 0.97 | 0.12 | 2.30 | 0.34 | 76.28 | 13.11 | 27.40 | 1.96 | 102.57 | 13.97 | 168.64 | 34.94 |
| Insert season | | | | | | | | | | | | | |
| Winter | 315 | 0.94 | 0.20 | 2.14 | 0.50 | 70.54 | 20.83 | 27.75 | 2.27 | 96.72 | 21.78 | 153.73 | 51.70 |
| Spring | 288 | 0.98 | 0.07 | 2.23 | 0.22 | 77.60 | 6.48 | 28.16 | 2.24 | 103.69 | 5.26 | 169.30 | 20.11 |
| Summer | 282 | 0.99 | 0.07 | 2.20 | 0.19 | 76.55 | 6.90 | 27.87 | 2.33 | 103.40 | 4.90 | 167.23 | 19.99 |
| Autumn | 386 | 0.97 | 0.09 | 2.27 | 0.45 | 71.81 | 13.72 | 27.87 | 2.18 | 98.04 | 14.57 | 156.06 | 37.83 |
| Size | | | | | | | | | | | | | |
| Small | 628 | 0.95 | 0.13 | 2.19 | 0.44 | 74.49 | 16.19 | 26.29 | 1.65 | 98.08 | 16.53 | 158.00 | 41.15 |
| Large | 643 | 0.99 | 0.10 | 2.23 | 0.29 | 73.24 | 11.20 | 29.48 | 1.51 | 102.23 | 11.20 | 163.85 | 30.65 |
| Shielding[a] | | | | | | | | | | | | | |
| Yes | 961 | 0.97 | 0.09 | 2.24 | 0.35 | 73.71 | 13.57 | 27.95 | 2.26 | 99.98 | 13.79 | 160.48 | 35.72 |
| No | 161 | 0.97 | 0.15 | 2.14 | 0.38 | 73.55 | 14.94 | 27.93 | 2.30 | 100.34 | 15.25 | 160.19 | 37.17 |
| Distance to corridor[a] | | | | | | | | | | | | | |
| 1st pen | 153 | 0.99 | 0.09 | 2.11 | 0.28 | 70.19 | 11.89 | 27.73 | 2.10 | 97.34 | 13.00 | 146.89 | 31.34 |
| 2nd pen | 127 | 1.00 | 0.12 | 2.04 | 0.31 | 67.90 | 12.53 | 29.15 | 1.75 | 98.34 | 14.60 | 144.17 | 34.50 |
| 3rd pen | 128 | 1.00 | 0.10 | 2.07 | 0.22 | 68.51 | 11.98 | 29.41 | 1.94 | 99.01 | 13.54 | 146.04 | 32.01 |
| 4th pen | 152 | 1.04 | 0.12 | 2.09 | 0.33 | 68.81 | 11.75 | 29.66 | 1.79 | 101.76 | 13.80 | 152.38 | 34.20 |

[a] Descriptive statistics relating to shielding and distance from corridor is only calculated for the groups which were included when fitting the models.

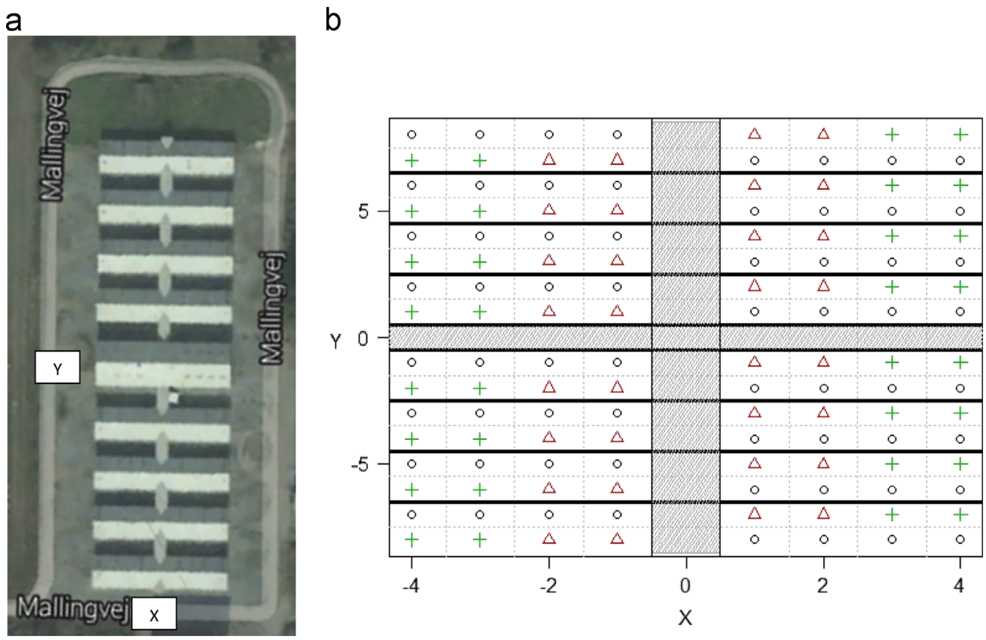



**Fig. 1.** (A) An aerial view photo of Bøgildgård. The symmetrical fish bone structure of the station is seen, with the different sections of the stations sticking out from the central corridor as white rectangle. (B) The (*X*,*Y*) coordinate system overlay. The vertical shaded area represents the central corridor. The horizontal shaded area represents the central section of the building which does not house pigs, separating the north- and south-facing ends of the station. There are eight sections on each side of the central corridor, as indicated by the solid lines. Each section is divided down the middle with four pens on each side, as indicated by the dotted lines. The thicker solid lines identify where wind shielding is placed between two neighboring sections. Dotted gray lines represent separate pens within a section. The color coded figures in the pens represent the breed of the groups which are most commonly found in a given pen. Circles: Duroc, Triangles: Danish Landrace, and Crosses: Yorkshire. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
*Source*: https://www.google.dk/maps/search/b%C3%B8gildg%C3%A5rd/@56.2746066,9.3889412,530m/data=!3m1!1e3

**Fig. 2.** The wooden walls separating shielding the space between the sections. The photo is taken on the East-side of the station.

contains 426 Duroc-, 267 Yorkshire- and 268 Danish Landrace-groups.

To determine the importance of wind shielding with respect to each of the three breeds, different models including this predictor were fitted to the data collected for the groups of each of the breeds. Interactions between shielding and each of the other two predictors were included. The start weight of the group was converted to a comparative start weight, *i.e.* the difference between the start weights of the group and the average start weight given the breed. A random effect of the pen was included in the model. Thus for each breed, a model with the following structure was fitted:

$$\text{Outcome} \sim \text{Comparative start weight} + \text{Insert season} + \text{Shielding} + (\text{Comparative start weight} * \text{Shielding}) + (\text{Insert season} * \text{Shielding}) + \text{Random (Pen)} + \text{Error} \quad (5)$$

*2.2.1.2. Effect of distance from corridor.* Because only Duroc groups are placed at all distances from the central corridor, only Duroc groups were included in the estimation of the significance of this factor. Furthermore, there is a tendency at Bøgildgård towards putting larger pigs further away from the central corridor, and to account for this, the groups were split into two subsets (large and small), depending on their weight at insertion. The cut-off between large and small was set at the average start weight for the breed, *i.e.* 29 kg (Table 1). Further, a random effect of the pen was included in the model. Thus for each start weight category the following model was fitted:

$$\text{Outcome} \sim \text{Insert season} + \text{Distance} + (\text{Insert season} * \text{Distance}) + \text{Random (Pen)} + \text{Error} \quad (6)$$

To determine the significances in differences between each of the individual levels, Tukey's Honest Significance Difference test was performed, using the R function TukeyHSD. This function provides a pair wise comparison of each of the included parameters.

*2.2.1.3. Interaction between shielding and distance to corridor.* To determine if any significant interaction effects exist between shielding and distance to corridor, a third model was fitted to the same set of Duroc data, as described in relation to Eq. (5). The model was as follows:

$$\text{Outcome} \sim \text{Insert season} + \text{Distance} + \text{Shielding} + (\text{Distance} * \text{Shielding}) + \text{Random (Pen)} + \text{Error} \quad (7)$$

## 3. Results

### 3.1. Effect of shielding

Table 2 shows the estimated coefficients for each of the factors of the reduced model of Eq. (5) for ADG and FCR. For Duroc pigs, shielding alone is not a significant predictor of ADG ($p=0.43$) but does appear to be so for FCR, with an absence of shielding resulting in a higher FCR ($p=0.038$). Shielding does however seem to interact negatively with larger comparative start sizes, in relation to ADG ($p=0.0002$). Shielding could not be shown to have any effect on ADG of Yorkshire and Danish Landrace pigs, but not being shielded appears to have a lowering effect on the FCR of Danish Landrace ($p=0.032$).

### 3.2. Effect of corridor

The initial descriptive analysis indicated that only the distance from the central corridor was important for ADG, regardless of the East–West direction, with ADG showing a positive correlation to greater distance (data not shown). For this reason, the data has been grouped according to the numerical *X*-values, *i.e.* the distance from the central corridor regardless of the East–West orientation, for the purpose of all further analysis.

The estimated coefficients for each of the significant predictors of ADG and FCR found by backward elimination of Eq. (6) are seen in Table 3. It is seen that insert season is a significant factor in relation to ADG for both large (lower panel) and small (upper panel) Durocs, as would be expected from the previous results.

As seen from Table 3, the distance from the central corridor is highly significant ($p<0.001$) for ADG in pigs which start out relatively large, with these pigs growing faster when they are further away from the corridor, but it does not appear to be significant for pigs that start out small. The specific differences between the effects of each pair of distances on the large pigs can be identified with Tukey's HSD test. The result of this test is shown in Table 4.

The most significant difference is seen between groups placed at the first and fourth pens from the corridor with the groups with the greater distance to the corridor growing an average of 48 g more per day. Less, but still significant at the 95% confidence level, is the difference between groups in the second and the fourth pens from the corridor, with an average difference of 27 g of growth per day. For FCR, only the interaction between insert season and pen number was found to be significant (Tables 3, $p=0.02$) by ANOVA during the backwards

**Table 2**
Estimates of the coefficients for the reduced models predicting ADG and FCR for each of the three breeds. The original model was Eq. (5), where estimates appear for one outcome but not the other; this is due to the given factor only being significant in predicting one of these outcomes. Standard deviations are calculated from the standard errors of the reduced model.

| Breed | Variable | ADG | | | FCR | | |
|---|---|---|---|---|---|---|---|
| | | Estimate | S.D. | *p*-Value | Estimate | S.D. | *p*-Value |
| Duroc | Intercept | 1.002 | 0.18 | < 0.0001 | 2.101 | 0.52 | < 0.0001 |
| | Comp. start size | 0.011 | 0.04 | < 0.0001 | 0.013 | 1.46 | 0.023 |
| | Insert season | | | 0.001 | | | 0.34 |
| | Autumn | 0 | | | 0 | | |
| | Spring | 0.012 | 0.13 | | 0.015 | 0.39 | |
| | Summer | 0.030 | 0.13 | | −0.023 | 0.38 | |
| | Winter | −0.006 | 0.11 | | −0.007 | 0.43 | |
| | Shielding | | | 0.43 | | | 0.038 |
| | Yes | 0 | | | 0 | | |
| | No | −0.023 | 0.19 | | 0.027 | 0.57 | |
| | Comp. start size: shielding | | | 0.0002 | | | |
| | Yes | 0 | | | | | |
| | No | 0.022 | 0.05 | | | | |
| | Insert season: shielding | | | 0.007 | | | 0.015 |
| | Autumn: Yes | 0 | | | 0 | | |
| | Spring: No | −0.009 | 0.14 | | −0.076 | 0.41 | |
| | Summer: No | −0.009 | 0.13 | | −0.022 | 0.4 | |
| | Winter: No | −0.098 | 0.14 | | −0.269 | 0.42 | |
| | Random pen effect | 0.000 | 0.09 | | 0.000 | 0.27 | |
| Yorkshire | Intercept | 1.002 | 0.14 | < 0.0001 | 2.097 | 0.18 | < 0.0001 |
| | Comp. start size | 0.905 | 0.09 | 0.005 | | | |
| | Insert season | | | | | | 0.007 |
| | Autumn | | | | 0 | | |
| | Spring | | | | −0.133 | 0.59 | |
| | Summer | | | | −0.191 | 0.58 | |
| | Winter | | | | −0.218 | 0.6 | |
| | Random pen effect | 0.000 | 0.09 | | 0.000 | 0.44 | |
| Danish Landrace | Intercept | 1.002 | 0.01 | < 0.0001 | 2.097 | 0.01 | < 0.0001 |
| | Comp. start size | 0.972 | 0.01 | 0.047 | 0.016 | 0.02 | 0.044 |
| | Insert season | | | 0.026 | | | |
| | Autumn | 0 | | | | | |
| | Spring | 0.012 | 0.02 | | | | |
| | Summer | 0.018 | 0.02 | | | | |
| | Winter | −0.025 | 0.01 | | | | |
| | Shielding | | | | | | 0.032 |
| | Yes | | | | 0 | | |
| | No | | | | −0.12 | 0.02 | |
| | Random pen effect | 0.000 | 0.09 | | 0.000 | 0.30 | |

elimination, and this was only for the pigs with the lower start weight. The Tukey HSD test showed that only eight of the 120 comparisons of combinations of pen number and insert season had significant differences ($p < 0.05$, data not shown). Given the number of comparisons, six significant differences would on average be expected by random chance at this level of significance.

### 3.3. Interaction between shielding and corridor

When reducing Eq. (7) via backwards elimination, the interaction- and shielding-factors were consistently removed, regardless of whether the model was fitted on all included Durocs simultaneously or on large and small starters separately (data not shown). This resulted in models which where, in effect, identical to those reached by reduction of Eq. (6).

## 4. Discussion

### 4.1. Shielding

Shielding on its own does not significantly affect the ADG of Duroc pigs, but there is a strongly significant positive correlation between shielding and the comparative start size in terms of ADG. We hypothesis that this effect is caused by the fact that larger pigs will produce larger amounts of body heat (Van Milgen and Noblet, 2003) which will be easier for the pigs to loose in a section which is not shielded against the wind. The interaction between shielding and season is also seen to be strongly significant in relation to ADG. A TukeyHSD analysis showed that five of the 28 shielding-season comparisons (14%) showed significant differences in growth rate. These were all the cases where being unshielded during winter was compared to being either shielded or unshielded during any other season. Pigs which are unshielded during winter consistently grow more slowly than the alternative

**Table 3**
The coefficients estimated for the factors included in the reduced model for ADG and FCR. The original model was Eq. (6), where estimates appear for one outcome but not the other; this is due to the given factor only being significant in predicting one of these outcomes. Standard deviations are calculated from the standard errors of the reduced model.

| Size | Variable | ADG | | | FCR | | |
|---|---|---|---|---|---|---|---|
| | | Estimate | S.D. | *p*-value | Estimate | S.D. | *p*-value |
| Small | Intercept | 1.000 | 0.25 | < 0.0001 | 2.11 | 0.99 | < 0.0001 |
| | Insert season | | | 0.0002 | | | 0.14 |
| | Autumn | 0 | | | 0 | | |
| | Spring | −0.003 | 0.19 | | −0.036 | 0.72 | |
| | Summer | 0.010 | 0.18 | | −0.084 | -0.69 | |
| | Winter | −0.082 | 0.19 | | −0.006 | 0.71 | |
| | Distance to corridor | | | | | | 0.31 |
| | 1st | | | | 0 | | |
| | 2nd | | | | −0.039 | 0.77 | |
| | 3rd | | | | −0.144 | 0.81 | |
| | 4th | | | | −0.073 | 0.77 | |
| | Insert season: distance | | | | | | 0.02 |
| | Autumn: 1st | | | | 0 | | |
| | Spring: 2nd | | | | 0.008 | 0.56 | |
| | Summer: 2nd | | | | 0.148 | 0.58 | |
| | Winter: 2nd | | | | −0.466 | 0.53 | |
| | Spring: 3rd | | | | 0.164 | 0.61 | |
| | Summer: 3rd | | | | 0.201 | 0.55 | |
| | Winter: 3rd | | | | 0.039 | 0.54 | |
| | Spring: 4th | | | | 0.041 | 0.55 | |
| | Summer: 4th | | | | 0.1 | 0.50 | |
| | Winter: 4th | | | | −0.064 | 0.60 | |
| | Random pen effect | 0.000 | | 0.13 | 0.060 | 0.32 | |
| Large | Intercept | 0.994 | 0.25 | < 0.0001 | | | |
| | Insert season | | | 0.033 | | | |
| | Autumn | 0 | | | | | |
| | Spring | 0.013 | 0.08 | | | | |
| | Summer | 0.028 | 0.08 | | | | |
| | Winter | 0.016 | 0.08 | | | | |
| | Distance to corridor | | | 0.0004 | | | |
| | 1st | 0 | | | | | |
| | 2nd | 0.021 | 0.10 | | | | |
| | 3rd | 0.023 | 0.10 | | | | |
| | 4th | 0.048 | 0.11 | | | | |
| | Random pen effect | 0.000 | 0.06 | | | | |

**Table 4**
Pairwise comparisons of the difference in ADG given each pair of distances from the central corridor.

| Comparison | Difference | Lower estimate | Upper estimate | Adjusted *p*-value |
|---|---|---|---|---|
| 2nd–1st | 0.021 | −0.009 | 0.053 | 0.253 |
| 3rd–1st | 0.023 | −0.007 | 0.054 | 0.198 |
| 4th–1st | 0.048 | 0.020 | 0.078 | 0.0001 |
| 3rd–2nd | 0.001 | −0.026 | 0.028 | 1.000 |
| 4th–2nd | 0.027 | 0.002 | 0.052 | 0.033 |
| 4th–3rd | 0.026 | 0.001 | 0.050 | 0.033 |

scenarios. This is consistent with the findings of Fitzgerald et al. (2009), as mentioned in the Section 1. Nonetheless, it is apparently at odds with the observation that FCR is significantly lowered when there is no shielding, most significantly during the winter. However, the average daily feed intake (ADFI) of unshielded pigs during winter was consistently significantly ($p \leq 0.01$) less than both shielded and unshielded pigs during any other season, with ADFI reductions ranging from 26 to 43 g, depending on the season. As seen from Eq. (2), this tendency to not eat would cause the FCR to be lower, making it superficially appear that the pigs are utilizing their feed more efficiently, but with no advantages in terms of growth. These interaction effects of shielding and seasons of the year could not be shown for Yorkshire or Danish Landrace. However, given the much smaller number of pigs of these breeds, compared to the number of Durocs, it is likely that the power of the study was simply not high enough to detect any effects on these breeds. Danish Landrace pigs do however seem to have a lower FCR when they are unshielded, regardless of size or insert season. To determine if this difference is practically significant, we calculated the amount of feed a group of shielded and unshielded Danish Landrace pigs would need to grow 70 kg, all other things being equal. This is simply calculated as weight gain multiplied by the FCR. The difference was found to be 12 g (146.79 kg and 146.67 kg for shielded and unshielded, respectively), which cannot be considered to be of any practical significance. This is also in line with the observation that no significant different difference was seen in ADFI for shielded vs. non-shielded Danish Landrace pigs (data not shown).

The negative coefficients, as seen in Table 2, between ADG and winter and the more positive coefficients between

ADG and summer suggest a positive correlation between warmer weather and faster growth, regardless of shielding. Negative coefficients are seen between FCR and all non-Autumn seasons, and so do not indicate an obvious relationship between warm or cold weather and FCR.

### 4.2. Corridor

A clear trend towards better growth at greater distance from the central corridor was observed in Duroc pigs with a start weight above average (29 kg). Large pigs at the far end of the sections were growing on average 48 g more per day. Given the mean ADG values shown in Table 3, and assuming an average start weight of 30 kg, a group of pigs in the first pen from the corridor would be expected to reach an average weight of 100 kg after 70 days. For a group in the fourth pen, this weight would be reached after only 67 days. This difference of three days means three days extra where the farmer has to spend feed and space on an old group rather than inserting a new group of growers. This trend was only observed for the larger pigs regardless of whether the sections were pointing towards the east or the west. As seen in Table 1, pigs that start out small generally also grow more slowly than larger pigs, and it thus might be that the effect of being small is greater than the effect of placement relative to the corridor, leading to the effect only being relevant in larger pigs. It thus makes sense to place the larger pigs away from the central corridor, whenever possible. This difference may possibly come about through differences in such factors as temperature, humidity and/or draft, which would be expected to be different next to a non-tight door to a long corridor with openings in both ends. The hypothesis that placements have shown effect on ADG through such temperature differences seems like a good potential explanation, given that temperature is well known to have great influence on the growth of pigs (Verstegen et al., 1978). Pigs placed by the door to the central corridor might also, for the same reasons, experience greater diurnal differences in temperature and humidity, which was shown by Lopez et al. (1991) to negatively influence the growth of pigs, compared to experiencing more constant temperature and humidity. Regarding the feed conversion ratio, a statistically significant effect was only seen for the interaction between the distance from the corridor and the insert season, and only for the smaller pigs. Since only 8 of the 120 permutation comparisons (6.7%) showed a significant difference at the 95% confidence level, given the Tukey HSD test, combined with the large standard errors observed for these interactions, we find it unwarranted to conclude that this is a real effect. This of course means that the observed effect of distance from the corridor on ADG does not come into effect through the efficiency by which the pig can utilize the feed for growth, but through the ADFI. Table 1 shows that pigs in the first pen generally eat less than those in the in the fourth pen during their time in the grower/finisher station (146.89 kg vs. 152.38 respectively). When focusing only on pigs that are larger than average at insertion, this difference becomes even more pronounced, with the average total consumption being 146.68 kg and 155.34 kg for the first and fourth pens, respectively. This is equivalent to an ADFI of 2.17 in the first pen and 2.26 kg/day in the fourth pen, a difference which is statistically significant ($p=0.042$). The hypothesis of this reduced ADFI being a result of variations in temperature and humidity is consistent with the findings of Huynh et al. (2005), who showed that temperature and humidity can drastically affect the voluntary feed consumption of growing pigs.

### 4.3. Perspectives

This study was performed with data from a pig research and testing station, but it stands to reason that the negative effect of growth by the proximity to the entrance of a given section would be relevant for pig producers in general. Whether or not the effect of this proximity is simply due to (diurnal) differences in temperature humidity and draft or some other factors, *e.g.* stress induced by human noises in the corridor, is not obvious from the data analyzed in this study, since no pen level temperature or humidity data was available. The temperature, humidity and draft hypotheses are however consistent with existing literature. If the case truly is that the differences in these factors, given placements and season, are affecting the ADG in the way we observe here, then it should be possible to simply measure these factors in the individual pens and fit a model to describe the ADG with an accuracy equal to or better than the models presented here. A relatively simple study could thus be done to support or dismiss this hypothesis. From these observations, we can however say that several sensors per section in any pig production station might be necessary in order to effectively monitor a pig production station, as significant differences are clearly found within a given section of such a station. At present the standard is at most one temperature sensor per section, which is also what is used at Bøgildgård, from where the data used in this study was obtained. A multitude of sensors would be needed for a complete monitoring of any herd, insuring a more uniform production.

Our findings of the effects of shielding in interaction with season on Duroc pigs suggest that larger pigs are generally better placed in unshielded pens, except during winter, where the lack of shielding shows a clear detrimental effect to the growth rate. It seems intuitively unlikely that this would only apply to Duroc and not the other breeds, which were included in this study. To determine whether this is in fact the case or not, a study including larger populations of non-Durocs would be needed. However, given that the effects of shielding were found to be completely and consistently insignificant when distance to the corridor and shielding were both included in the same model suggest that shielding should at most be a secondary concern when placing new growers in the station.

## 5. Conclusion

Larger Duroc pigs are found to generally grow faster when placed in sections without shielding. An important exception is when the pigs are inserted in the grower/finisher station during the winter, in which case lack of shielding has an overwhelmingly negative effect on Duroc growth rate. No effect of shielding could be demonstrated for Yorkshire or Danish Landrace. Greater pen distance from the central corridor of the station is found to have a significant positive effect on the growth of Duroc boars with above average start weight. Neither shielding nor distance from the central corridor is shown to affect feed conversion ratio. Rather, the effect of greater growth further from the central corridor appears to come about through a different daily feed intake between groups of pigs close to and further away from the central corridor of the station.

## Conflict of interest statement

None.

## Acknowledgments

This research was carried out with support from The Danish Council for Strategic Research (The PigIT project, Grant no. 11-116191). The authors would further like to thank the Danish Pig Research Center (Danish: Videncenter for Svineproduction) for providing the data used in this study. Lastly the authors wish to thank the English tutor Colleen Golden Ussery for proofreading previous versions of this manuscript.

## References

Aarnink, A.J.A., et al., 2006. Temperature and body weight affect fouling of pig pens. J. Anim. Sci. 84 (8), 2224–2231.

AU Foulum. 2013. "AU Foulum." ⟨http://dca.au.dk/en/about_dca/au-foulum/⟩ (20.01.14).

Close, W.H., Heavens, R.P., Brown. D. 1981. The effects of ambient temperature and air movement on heat loss from the pig. Anim. Prod. 32(01), 75–84. ⟨http://journals.cambridge.org/abstract_S0003356100024806⟩ (20.01.14).

Danish Agriculture and Food Council, 2013. Pig Research Centre. ⟨http://www.pigresearchcentre.dk/⟩ (07.06.13).

Fitzgerald, R.F., et al., 2009. Factors associated with fatigued, injured, and dead pig frequency during transport and lairage at a commercial abattoir. J. Anim. Sci. 87 (3), 1156–1166. ⟨http://www.ncbi.nlm.nih.gov/pubmed/19028860⟩ (20.01.14).

Huynh, T.T.T., et al., 2005. Effects of increasing temperatures on physiological changes in pigs at different relative humidities the online version of this article, along with updated information and services, is located on the world wide web at: effects of increasing temperature. J. Anim. Sci. 83, 1385–1396.

Lopez, J., Jesse, G.W., Becker, B.A., Ellersieck, M.R., 1991. Effects of temperature on the performance of finishing swine: II. Effects of a cold, diurnal temperature on average daily gain, feed intake, and feed efficiency. J. Anim. Sci. 69, 1850–1855.

Mount, L.E., 1966. The effect of wind-speed on heat production in the new-born pig. Q. J. Exp. Physiol. 51, 18–26.

Verstegen, M.W.A., Brascamp, E.W., vander Hel, W., 1978. Growing and fattening of pigs in relation to housing and feeding level. Can. J. Anim. Sci. 58, 1–13.

Van Milgen, J., Noblet, J., 2003. Partitioning of energy intake to heat, protein, and fat in growing pigs. J. Anim. Sci. 81, 86–93.

Segalés, Joaquim, et al., 2012. Exploratory study on the influence of climatological parameters on mycoplasma hyopneumoniae infection dynamics. Int. J. Biometeorol. 56 (6), 1167–1171. ⟨http://www.ncbi.nlm.nih.gov/pubmed/21904808⟩ (20.01.14).

The R Core Team, 2013. 1 *R:* A Language and Environment for Statistical Computing. 2.15.3 ed. R Foundation for Statistical Computing.

Wilson, Stephen, et al., 2012. Vaccination of piglets at 1 week of age with an inactivated mycoplasma hyopneumoniae vaccine reduces lung lesions and improves average daily gain in body weight. Vaccine 30 (52), 7625–7629. ⟨http://www.ncbi.nlm.nih.gov/pubmed/23084853⟩ (20.01.14).

Young, B.A., 1981. Cold stress as it affects animal production. J. Anim. Sci. 52, 154–163. ⟨http://www.ncbi.nlm.nih.gov/pubmed/7240034⟩ (12.09.13).

# Chapter 6: Paper 2

Published in *Livestock Science*:

Jensen, D.B. & Kristensen, A.R., 2016. Temperature as a predictor of fouling and diarrhea in slaughter pigs. *Livestock Science*, 183, pp.1–3.

Livestock Science 183 (2016) 1–3

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Short communication

# Temperature as a predictor of fouling and diarrhea in slaughter pigs

Dan Børge Jensen*, Anders Ringgaard Kristensen

*Department of Large Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 2, DK-1870 Frederiksberg C, Denmark*

ARTICLE INFO

*Article history:*
Received 24 June 2015
Received in revised form
6 November 2015
Accepted 8 November 2015

*Keywords:*
Diarrhea
Model
Pen fouling
Porcine
Prediction
Slaughter pig

ABSTRACT

The PigIT Project aims at improving welfare and productivity of slaughter pigs by integration of various sensor systems for alarm purposes. Here we present an exploratory analysis to assess the predictive value of temperature sensor data with respect to pen fouling and diarrhea. We recorded the temperature at two locations in 8 pens between November 2013 and December 2014. A single logistic regression model was made to express the probability of either diarrhea or fouling per pen per day, and was reduced via backwards elimination. The predictive performances were evaluated by the area under the receiver operating characteristics curve (AUC). Indiscriminant prediction of either event reached an AUC of 0.80. Similar performances were seen when predicting each of the events on their own using the same model, with AUC values at 0.78 and 0.81 for diarrhea and fouling, respectively. Thus, temperature information seems to provide predictive value in relation to fouling and diarrhea. It would be meaningful to combine this information with other available data by using more advanced models to achieve an optimal predictive power.



## 1. Introduction

General and political interest in production and animal welfare is currently at an all time high in Denmark and other western countries. Denmark alone produces nearly 30 million pigs annually, distributed between just over 3000 farms (Landbrug og Fødevarer, 2014). Large herds are at increased risk of infectious diseases (Claes et al., 2002), and infectious disease will be more likely to persist for larger herds (Evans et al., 2010). It is further known that pig health and productivity is affected by a range of stress factors. One such factor is the temperature, where especially diurnal changes can cause a stress response, resulting in slower growth and higher feed intake (Lopez et al., 1991). Temperature is further known as a key factor for the onset of pen fouling (Aarnink et al., 2006), where the pigs will rest in the excretion area and in return excrete in the resting area.

Here, we wish to evaluate the potential of pen level temperature measurements for predicting pen level outbreaks of two undesired events in pig production, namely diarrhea and pen fouling. An effective prediction of such undesired events would allow the farmer to react proactively to a problem, thus improving the overall health and welfare of the herd, and in return secure a higher production for the farmer.

## 2. Materials and methods

### 2.1. Data source

The data used for this study were collected for the PigIT Project[1] in the finisher unit of a commercial Danish pig farm. Temperature data was collected continuously in 8 pens in two separate sections. Each pen contained 18 pigs at insertion, sorted by sex and size. Two neighboring pens (a double-pen) always shared feed and water supply. Data from two such double-pens were included from each of the two sections. Data were collected between November 20th 2013 and December 12th 2014, during which time three new batches of 30 kg pigs were inserted in each pen.

The climate was controlled at section level by a climate computer, model Dol 234 from the company Skov A/S. This computer adjusted the climate via a combi-diffuse ventilation system, sprinklers above each pen and heating pipes installed in the back walls. The set points (depending on age of pigs) were 15–20 °C for temperature and 70–75% for humidity. Heating was activated when the section temperature was 1 °C below the set point. The sprinklers could be activated between 9 AM and 8 PM if temperatures were above the set point. Section temperatures of 0.5 °C above the set point would activate the sprinklers at 0.1% capacity

* Corresponding author.
*E-mail address:* daj@sund.ku.dk (D.B. Jensen).

[1] http://pigit.ku.dk/.

http://dx.doi.org/10.1016/j.livsci.2015.11.007
1871-1413/

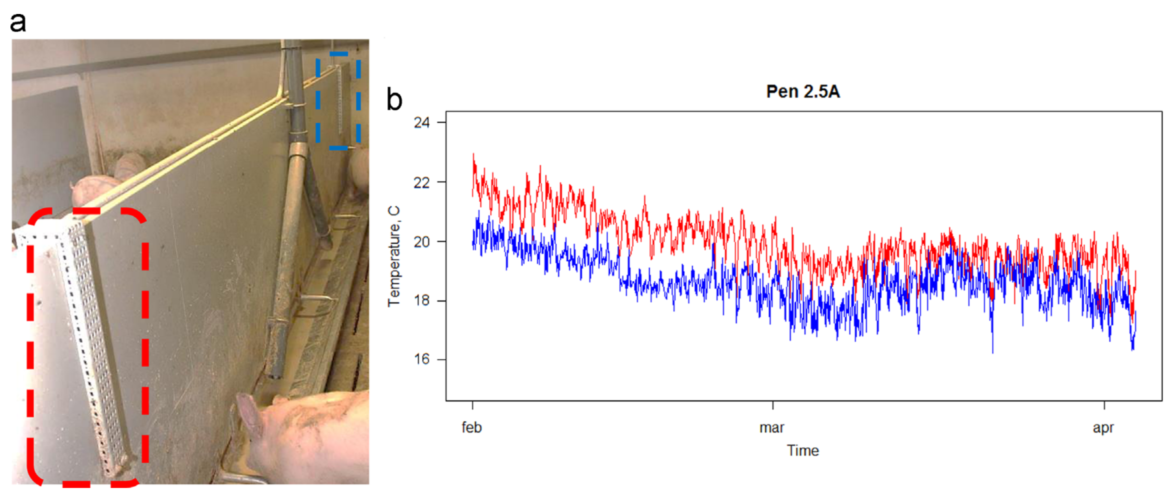



**Fig. 1.** a) The pen set-up, with two thermometers installed per pen. One is by the drinking nipple and resting area (smaller dashed rectangle), the other is by the corridor and excretion area (larger dashed rectangle). b) An example of the temperature measurements from a single pen. Temperatures at the drinking nipple are generally a few degrees warmer than temperatures near the corridor.

for 0.5 min with 60 min intervals. Section temperatures 3 °C or more above the set point would activate the sprinklers for 1.5 min with intervals of 30 min. For temperatures between 0.5 and 3 °C above the set point the capacity, frequency and duration of sprinkling are interpolated. Ventilation was controlled according to a curve, starting at 7 $m^3$/pig at insertion and going up to 15 $m^3$/pig at the time of delivery.

The pen level temperature data used in this study were collected automatically by two thermometers in each pen, as seen in Fig. 1a. Specifically VE10-A temperature sensors from the company VENG System were used. The thermometers were placed by the drinking nipple (small dashed square) and by the corridor (larger dashed square). The temperatures at these two locations were consistently distinct, with the temperature by the drinking nipple generally being a few degrees warmer than at the corridor, as exemplified in Fig. 1b.

### 2.2. Modeling

All data management and modeling was done using R (The R Core Team, 2013). The continuous temperature data were averaged over periods of 60 min. It was further summarized to a daily level by finding the highest and lowest averaged temperature as well as the greatest decrease and increase in temperature. The data were subsequently split into a training set (13 events total) and a test set (15 events total). In each of the two sections, all data from one double-pen were assigned to the training set, while all observations of the other double-pen were assigned to the test set. A logistic regression model with no interaction effects, describing the daily probability of any event (diarrhea or pen fouling), was fitted to the learning set data using the built-in function glm. The model was subsequently reduced by backwards elimination, using the built-in function step. The reduced models were used to predict the undesired events in the test set, when the probability of the events surpassed a set threshold. A positive prediction was considered a true positive if it was made at most 3 days before or 1 day after an event observation. The predictive performances were evaluated using area under the receiver operating curve (AUC) (Zweig and Campbell, 1993), which was calculated using the function auc from the MESS package (Ekstrom, 2013).

## 3. Results and discussion

Table 1 summarizes the logistic regression model for predicting any of the two events, when reduced to including only significant

**Table 1**
The variables included in the reduced logistic regression model after backwards elimination, along with parameter estimates, standard errors and *p*-values.

| Variables | Estimate | Std. error | *p*-Value |
|---|---|---|---|
| (Intercept) | −12.78 | 3.53 | 0.0003 |
| Maximum temperature, drinking nipple | −0.89 | 0.35 | 0.011 |
| Minimum temperature, drinking nipple | 1.24 | 0.43 | 0.004 |
| Greatest temperature decrease, drinking nipple | −2.13 | 0.87 | 0.014 |
| Greatest temperature increase, corridor | 2.70 | 0.90 | 0.003 |
| Greatest temperature decrease, corridor | 1.53 | 0.89 | 0.086 |

or borderline significant variables. It is seen that high rates of both temperature increase and decrease, measured near the corridor, is associated with a higher risk of undesired events. This could indicate that the pigs are generally sensitive to sudden changes in temperature, which concievably could cause them to become stressed (Lopez et al., 1991). Interestingly, a high rate of temperature decrease near the water nipple is associated with a reduced risk of undesired events. This makes sense, given that such quick reductions in temperature is what would be seen when the sprinkler system is activated in response to undesirably high temperatures, which is a known causal factor for e.g. pen fouling. This is similarly reflected by the fact that higher values of the lowest temperature, recorded during a 24 h period, is positively associated with higher risk of undesired events overall.

Fig. 2 shows the receiver operating characteristics curves for prediction on the independent test set of any undesired events (empty circles), as well as pen fouling and diarrhea separately

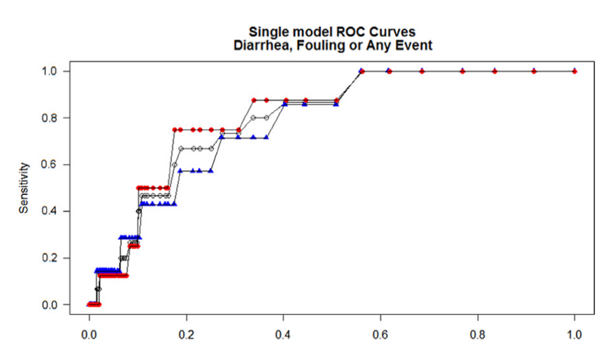



**Fig. 2.** ROC curves for any event (empty circles), pen fouling (solid circles), and diarrhea (triangles).

(solid circles and triangles, respectively). The corresponding AUC values are 0.80, 0.78 and 0.81 for any of the events, diarrhea, and fouling, respectively. As is seen, the predictive performance is consistently better than would be expected from random chance, proving that the variables included in the model described above contribute genuinely useful information regarding the risk of these undesired events. However, the trade-off between sensitivity and specificity is still apparent. For example, holding the sensitivity of the overall event prediction at 0.80, as has been described as necessary for when detecting e.g. mastitis in cows (Hogeveen et al., 2010) would result in a false positive rate of 34%. It should however be noted that this shortcoming could well be a consequence of the relatively simple linear decision boundary used by the logistic regression model. It could well be that an even better performance would be seen with the non-linear decision boundaries used by e.g. artificial neural networks. However, comparing the performance of such methods to the method applied here is beyond the scope of this paper.

## 4. Conclusion

It is shown that temperature data recorded at the pen level contains information, which is applicable to prediction of pen fouling and diarrhea up to 3 days before these events occur. The area under the receiver operating characteristics curves for indiscriminant, diarrhea, and fouling predictions are 0.80, 0.78 and 0.81, respectively. However, the logistic regression method used in this study is not likely to be the best method for practical purposes, as the achieved trade-off between sensitivity and specificity is such that this information is not likely to be practically useful on its own. Other methods with non-linear decision boundaries should thus be tested in the future.

## 5. Future scope

In the PigIT project, we are currently collecting data on water and feed consumption, live weight and section humidity as well as pen level temperature, and there are plans for including automatic monitoring of pig activity. The information contained in this data could conceivably be combined with the temperature data presented in this paper, using a number of methods, such as multivariate dynamic linear modeling (West and Harrison, 1997), (naïve) Bayesian networks, artificial neural networks or some combination of these and other methods. We expect such information integration to yield better predictions than any one line of evidence can provide on its own.

## Acknowledgments

This research was carried out with support from the Danish Council for Strategic Research (The PigIT Project, Grant number 11-116191). We further wish to thank the anonymous farmer, the Danish Pig Research Center and the technical staff at Aarhus University for installing and supervising the sensors and for taking care of the daily observations in the herd.

## References

Aarnink, A.J.A., Schrama, J.W., Heetkamp, M.J.W., Stefanowska, J., Huynh, T.T.T., 2006. Temperature and body weight affect fouling of pig pens. J. Anim. Sci. 84, 2224–2231. http://dx.doi.org/10.2527/jas.2005-521.

Claes, E., Mousing, J., Schirmer, A.L., Willeberg, P., 2002. Infectious and rearing-system related risk factors for chronic pleuritis in slaughter pigs. Prev. Vet. Med. 54, 337–349. http://dx.doi.org/10.1016/S0167-5877(02)00029-6.

Ekstrom, C., 2013. MESS | inside-R | A Community Site for R [WWW Document]. URL ⟨http://www.inside-r.org/packages/cran/MESS⟩ (accessed 11.02.15).

Evans, C.M., Medley, G.F., Creasey, S.J., Green, L.E., 2010. A stochastic mathematical model of the within-herd transmission dynamics of porcine reproductive and respiratory syndrome virus (PRRSV): fade-out and persistence. Prev. Vet. Med. 93, 248–257. http://dx.doi.org/10.1016/j.prevetmed.2009.11.001.

Hogeveen, H., Kamphuis, C., Steeneveld, W., Mollenhorst, H., 2010. Sensors and clinical mastitis–the quest for the perfect alert. Sensors 10, 7991–8009. http://dx.doi.org/10.3390/s100907991.

Landbrug og Fødevarer, 2014. Statistik 2013, svinekød.

Lopez, J., Jesse, G.W., Becker, B.A., Ellersieck, M.R., 1991. Effects of temperature on the performance of finishing swine : II. Effects of a cold, diurnal temperature on average daily gain, feed intake, and feed efficiency. J. Anim. Sci. 69, 1850–1855.

The R Core Team, 2013. R : A language and environment for statistical computing, 2.15.3 ed. R Foundation for Statistical Computing.

West, M., Harrison, J., 1997. Bayesian Forecasting and Dynamic Models, 2nd ed. Springer, New York, USA.

Zweig, M.H., Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. 39, 561–577.

# Chapter 7: Paper 3

Published in *Proceedings for the EC-PLF conference in Milan, 15-18 September 2015*

Jensen, D.B. et al., 2015. A multi-dimensional dynamic linear model for monitoring slaughter pig production. In *7th European Conference on Precision Livestock Farming*. pp. 503–512.

# A multi-dimensional dynamic linear model for monitoring slaughter pig production

Dan Børge Jensen[1,*], Cécile Cornou[1], Nils Toft[2], Anders Ringgaard Kristensen[1]

[1]*Department of Large Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 2, DK-1870 Frederiksberg C, Denmark*
[2]*National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark*

*daj@sund.ku.dk

## Abstract
Scientists and farmers still lack an efficient way to unify the large number of different types of data series, which are increasingly being generated in relation to automatic herd monitoring. Such a unifying model should be able to account for the correlations between the various types of data, resulting in a model which could potentially yield more information than can be gained from the individual components separately. Here we present such a model for monitoring slaughter pig production, in the form of a multivariate dynamic linear model. This model unifies three types of data (live weight, feed- and water consumption), measured at different levels of detail (individual pig and double-pen level) and with different observational frequencies (weekly and daily), using series collected for the Danish PigIT project. The presented three-dimensional model serves as a proof of concept, and it should be straightforward to expand it with additional data types.

## Key words
Dynamic linear model, information unification, modeling, monitoring, pig production

## Introduction
For many years, a whole range of sensors have been available for monitoring variables relevant for *e.g.* mastitis detection in dairy cows (Viguier *et al.* 2009) and the application of sensor technology is slowly being introduced for pig herd monitoring. The idea is that the collected data, combined with a proper alarm system, can provide the farmer with early alarms, thus allowing proper proactive responses to undesired changes in the herd. However, there are serious issues with the methods currently described. First, the trade-offs of sensitivity and specificity are generally unacceptable (Hogeveen *et al.* 2010). Furthermore, the systems tend to consider each monitored variable in isolation, so that an alarm is based on the value of just one variable, and no interaction effects are considered. We therefore believe that better integration of the available information could yield better methods for prediction of animal health states.
We suggest employing a multivariate dynamic linear model (DLM) (West & Harrison 1997) as a means of obtaining a more holistic monitoring of animal herds, taking into account the interconnectedness of all the variables of interest. Univariate DLM's have previously been

attempted for automated estrus detection in sows (Ostersen *et al.* 2010) and a multivariate DLM has been used to predict litter sizes in sows (Bono *et al.* 2012), but where only one type of information was considered. In general the use of DLM is under-utilized in the fields of animal- and veterinary science.

This paper serves as a proof of concept, demonstrating the use of a multivariate DLM for monitoring slaughter pigs in a Danish finisher unit, in terms of live weight, feed usage and water usage.

## Materials and Methods

### Data source

The work described in this paper was done using data collected for the PigIT Project[1] in a commercial Danish pig farm. Specifically, the data were collected in the farm's finisher unit, housing slaughter pigs while they grow from roughly 30-100 kg. The unit consists of five sections, each with 14 pens. Each pen contains 18 pigs (at insertion), sorted by sex and size. The climate within each section is controlled by a combi-diffuse ventilation system, computer-controlled sprinklers above each pen and heating pipes installed in the back walls.

For the PigIT Project, a number of sensors have been installed to automatically record data on feed usage, water flow to the drinking nipples and temperature in 16 of the 70 pens in the finisher unit.

Liquid feed is dispensed automatically into troughs, shared between two neighboring pens, as seen on Figure 1 A. Two such feed-sharing pens will be referred to as a *double pen*. The expected amount of liquid feed required for a given double pen is adjusted regularly by manual observation of how much of the dispensed feed has been left uneaten.
Water is dispensed from drinking nipples which, like the feed dispensers, are also shared between two pens in a double pen, as seen in Figure 1 B. In the 16 PigIT-pens (eight double pens) flow meters are installed above the water nipple to measure water flow to the double pen.

In addition, the individual pigs from two double pens in the same section (section 2) are manually weighed once per week from insertion until the first pigs from that section are sent to the abattoir. The weight measurements are performed with the pig scale depicted on Figure 1 C. The pigs in these pens are individually identified with RFID ear tags.

The data applied in this paper were collected from those two double pens, where live weight was recorded in addition to feed consumption and water flow. The used data were collected between November 20th 2013 and December 12th 2014, and included four separate insertions of new pigs. Thus the following models were based on a total of eight separate sets of observations.

[1] http://pigit.ku.dk/

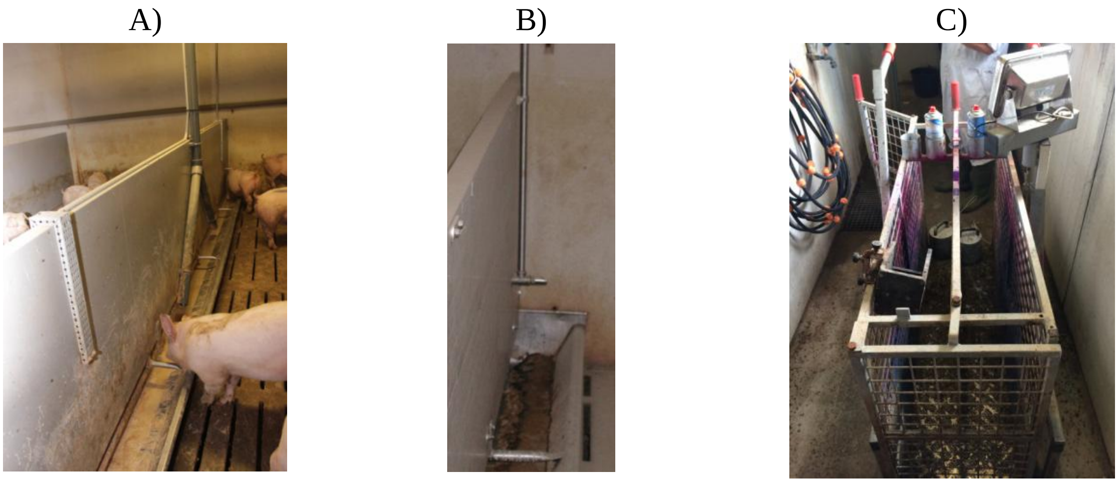



**Figure 1: The sources of the data used in this paper. A) Liquid feed dispensed to the double pen by the feeding system. B) Water consumed by the pigs in a double pen, recorded by a flow meter above drinking nipple. C) Scale for manual recording of individual animal weights.**

### Modeling

All analysis, modeling and representations of the data were done using the software R, a language and environment for statistical computing (The R Core Team 2013).

In this paper, we aim to demonstrate a method for meaningfully combining multiple different kinds of observational data (live weight, feed usage and water flow). From farm records on insertion and removal of individual pigs, it was possible to know how many pigs were in a given pen at any given time, and this information was used to normalize the feed usage and water flow to daily averages per pig in the double pen. Although the live weights of the pigs were recorded individually, these too were aggregated to a per pig average for the double pen, in order to simplify the model.

A multivariate dynamic linear model (DLM), as described by West and Harrison (1997), was the method chosen to combine the data. In general, a DLM consists of an observation equation and a system equation (Equations (1) and (2), respectively).

$$Y_t = F_t'\theta_t + v_t, \qquad v_t \sim N\left(\underline{0}, V\right) \tag{1}$$

$$\theta_t = G_t\theta_{t-1} + w_t, \qquad w_t \sim N\left(\underline{0}, W\right) \tag{2}$$

Equation (1) describes how the values of an observation vector ($Y_t$) depends on an unobservable parameter vector ($\theta_t$) to time $t$. The unit of time used in this model was one day.

In our case, the parameter vector contains the estimated underlying values for live weight ($LW$), feed usage ($Feed$) and water flow ($Water$), as well as the rates at which those same values change at time $t$ ($dLW$, $dFeed$, $dWater$, respectively), as seen in Table 1, $\theta_t$. The underlying values at time $t$ were estimated using a Kalman filter as described by West and Harrison (1997). In short, the Kalman filter is a method for filtering noise from the data by considering the actual observations, the error in the model forecasts and the systematic and observational variances.

For our purpose, the (transposed) design matrix has a structure with a basis as seen in Table 1 ( $F_t^{'}$). This structure serves to separate the estimated underlying values of the live weight, feed usage and water flow in the parameter vector from their respective trend values, in accordance with Equation (1). The structure is varied according to which variables are observed for a given time $t$, with the first, second and third row being included when live weight, feed usage and water flow are observed, respectively. Thus missing observations will be ignored when the parameter vector is updated.

The structure of the system matrix ($G_t$) is constant in our case (Table 1, $G_t$). This structure serves to add the trend values to the corresponding estimated values of the three parameters of interest, thus updating their values from time $t-1$ to time $t$, in accordance with Equation (2).

The initial values of live weight, feed usage and water flow were estimated from all available data as the average, normalized values observed on the first day of a batch insertion. The initial growth trend for live weight and feed usage were estimated as the average daily change in those values between the first and eighth day of observation. The water flow was seen to vary greatly from day to day, but did not follow any general trend over the grower/finisher periods. The initial rate of change was therefore set to 0.

**Table 1: The structures of the three matrices, presented in Equations (1) and (2), as they apply to the data used for this paper. $\theta_t$: The Parameter Vector. $F_t^{'}$: The Design Matrix (transposed). $G_t$: The System Matrix**

| $\theta_t$ | $F_t^{'}$ | $G_t$ |
|---|---|---|
| $\begin{bmatrix} LW_t \\ dLW_t \\ Feed_t \\ dFeed_t \\ Water_t \\ dWater_t \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ | $\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ |

The observational co-variance matrix ($V$) and the systematic co-variance matrix ($W$) were estimated from all available data, using the expectation maximization (EM) algorithm, as described by West and Harrison (1997), until convergence, which by visual inspection was found to occur after 50,000 iterations.

### Unification of model forecast errors

Once the parameters defining the DLM had been estimated, the average live weight, feed usage and water flow per pig in each of the eight separate batch observation series were modeled. During the modeling, a vector $(e_t)$ of forecast errors (*Observed values – Forecasted values*) was continuously generated for each time step. In addition, a matrix describing the forecast co-variances $(Q_t)$ was continuously generated, as described by West and Harrison (1997). A Cholesky decomposition was calculated for $Q_t$ using the R function chol. Using the decomposed matrix $\left(C_{Qt}\right)$, the error vector was transformed, as seen in Equation (3).

$$u_t = C_{Qt}^{-1} \cdot e_t \tag{3}$$

This transformation ensures that the transformed error values in $u_t$ are mutually independent and each follow a standard normal distribution. Thus a single value measuring the square of deviation from 0, the mean within this frame of reference, can be easily calculated for the set of forecast errors, as seen in Equation (4).

$$d_t^2 = u_t^{'} \cdot u_t \tag{4}$$

This unified error will follow a $\mathcal{X}^2$ distribution with $n$ degrees of freedom, where $n$ is the number of elements in $u_t$. Thus $d_t^2$ can be plotted to a conventional Shewhart control chart (Montgomery 2005) to allow for an easy monitoring of the complex system. The upper control limit was set to the 0.99 quantile of the $\mathcal{X}^2$ distribution. To allow for a constant control limit in response to varying degrees of freedom, $d_t^2$ was adjusted, according to Equation (5).

$$d_{adj.t}^2 = d_t^2 \cdot \left(\frac{\mathcal{X}^2(0.99, 3)}{\mathcal{X}^2(0.99, n)}\right) \tag{5}$$

## Results and discussion

### Model parameter values

The estimated initial values of the parameter vector are seen in Equation (6).

$$\theta_0 = (29.0\,, 0.65\,, 3.3\,, 0.79\,, 0.6\,, 0.0)^{'} \tag{6}$$

As is seen, the average pig initially weigh 29 kg, grow at a rate of 650 grams per day and eats 3.3 kg feed per day with a daily increase of 790 grams. The normalized water flow to the double pen is 0.6 liters per day per pig, with no (0.0) systematic daily change.

The matrices describing the observational and systematic co-variances are seen in Equations (7) and (8), respectively.

$$V = \begin{bmatrix} 2.37 & 5.22 & -3.49e-4 \\ & 10.43 & -2.23e-4 \\ & & 6.62e-5 \end{bmatrix} \tag{7}$$

$$W = \begin{bmatrix} 1.11 & 1.70e-3 & -1.41 & -2.80e-3 & 1.48e-2 & 7.37e-6 \\ & 2.61e-6 & -5.00e-4 & -3.53e-6 & -1.88e-5 & 0.00 \\ & & 4.52 & 1.37e-2 & -5.30e-2 & -3.31e-5 \\ & & & 5.92e-5 & -1.00e-4 & 2.53e-9 \\ & & & & 1.61e-1 & 2.17e-3 \\ & & & & & 1.61e-8 \end{bmatrix} \tag{8}$$

Notice that $V$ has a 3x3 structure, consistent with the three values which can be observed at each time $t$, while $W$ has a 6x6 structure, consistent with the six values in the parameter vector.
It is worth noting that the diagonal values in both matrices would have been the same if each variable of interest had been modeled separately. It is thus the co-variances outside the diagonals which provide the extra information about the interconnectedness of each of the monitored variables.

**Modeling**
The DLM defined as described in the previous sections was used to model each of the eight available sets of batch data. Figure 2 shows three notable examples of the output of this modeling. These are the batches inserted on July 7$^{th}$ 2014 to pen number 2.5 and 2.10 (top and bottom row, respectively) and the batch inserted on October 9$^{th}$ 2014 to pen number 2.10 (middle row).

The left column of Figure 2 shows the observed values for mean live weight (circles), feed usage (triangles) and water flow (solid squares). In addition, the left column shows the filtered mean, as estimated by the DLM, for live weight (solid line), feed usage (thick dashed line) and water flow (dotted line).

The right column shows the Shewhart control charts of the adjusted unified forecasts errors (circles connected by red lines), according to Equation (5). The horizontal lines in the control charts show the control value, *i.e.* the 0.99 quantile of the $\mathcal{X}^2$ distribution with 3 degrees of freedom (11.34).
For both columns, observations of diarrhea and pen fouling are marked by vertical lines. Diarrhea is marked by thick dashed lines, while solid lines represent pen fouling.

As is seen, the output in the top row is from a batch where no undesired events were observed, and all unified errors are all well below the control limit.

For the middle batch, pen fouling is observed twice (on the 3$^{rd}$ and 4$^{th}$ of December) and one case of diarrhea is observed around the 20$^{th}$ of November. The first pen fouling event falls just below the control line, but the second one coincides with a very clear spike in the unified error, which would yield a successful alarm. However, the system fails to raise an alarm about the diarrhea. This is probably because this event occurs during a period of time where water data is not available, which would be expected to strongly correlate with diarrhea.

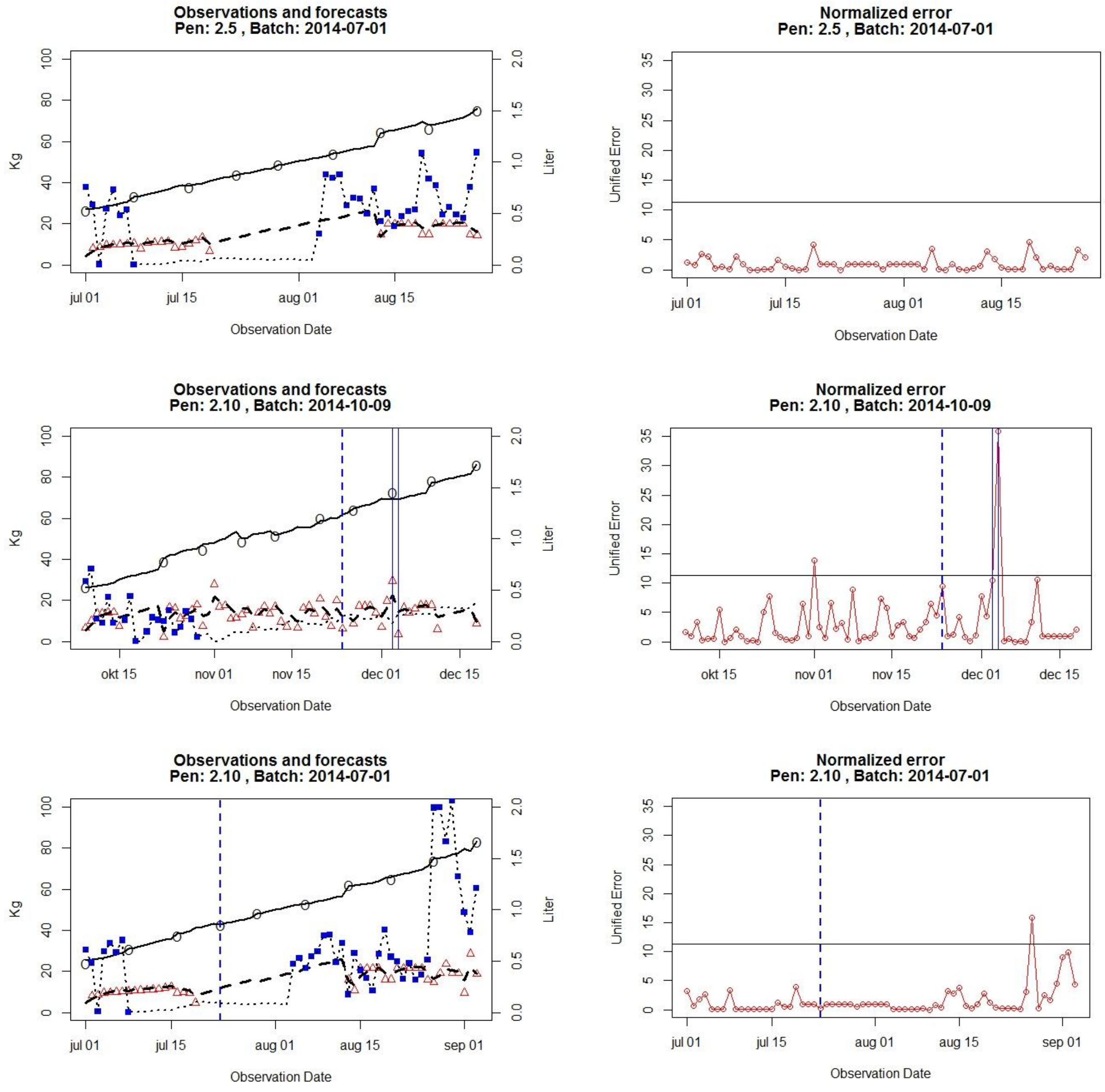



**Figure 2: Left column: the observed values of mean live weight (circles), feed usage (triangles) and water flow (solid squares) per pig in three separate batches. In addition, the filtered mean values, estimated by the model, for live weight (solid line), feed usage (dashed line) and water flow (dotted line). Right column: the unified forecast error for mean live weight, feed usage and water flow per pig, corresponding to the observations depicted in the left column. Vertical lines in both columns: observations of diarrhea (dashed) and pen fouling (solid).**

The bottom batch provides two interesting examples of how the system can fail in its function. First, an undesired event (diarrhea) is observed at July 20$^{th}$. This event happens to coincide with a relatively long period of time where the data on feed usage and water flow are both missing, and only the weight observation of that day is available. It can only be assumed that this information, especially regarding water flow, would have contributed to a more extreme unified error, and thus this example illustrates the value of having functioning sensors throughout a monitoring period.
Conversely, around the 25$^{th}$ of August, a tall peak is seen in the unified error, in spite of there being no observed undesired events. This peak is seen to be caused by a sudden dramatic and uncharacteristic increase in the water flow. From temperature records it can be found that this increase in water flow coincides with a sudden increase in temperature inside the double pen 2.10, while a similar temperature increase was not experienced in double pen 2.5 (data not shown).

**Perspectives**

When employing a dynamic linear model, an uncharacteristically large forecast error (here the limit was set to 11.34), is an indication that the observed system has changed significantly from the assumptions of the model. Thus, if the model has been optimized for describing a perfectly healthy batch of pigs, uncharacteristic forecast errors would likely indicate an outbreak of disease. To what extend the method demonstrated in this paper allow for more accurate disease detection, compared to other methods, will be a subject for further studies.

It should be noted that the form of forecast error unification demonstrated here can only yield an absolute magnitude of the error, and thus, unlike conventional univariate control charts, this control system cannot take into account whether some errors are positive and others negative. This potential problem could be circumvented by parsing the separate, non-unified errors to other classification systems, *e.g.* artificial neural networks or Bayesian classifiers. This would require separate training and validation of these systems, in addition to what is needed for the DLM itself. However, it is conceivable that such parsing could yield better detection of undesired events, and even allow for specific error patterns to be mapped to specific conditions, which would be another natural objective for further studies.

Whether or not a multivariate DLM defined from one herd can be directly applied in another herd or between different breeds of pigs, and how often such models need to be updated to keep up with the biological changes from breeding, are additional questions requiring further studies to answer.

Furthermore, here we have demonstrated the method with three measurable variables, but it would be trivial to adapt the model to include more (or fewer) lines of evidence, depending on data availability. We could envision modeling the live weight of each pig in the double pen individually or including the modeling of some measure of activity captured by video, etc. All that is needed is to design the appropriate design- and system matrices and the availability of the relevant data.

Lastly, this paper showcased the use of a multivariate DLM for monitoring slaughter pig production, but this method could just as well be employed in any animal production where data is routinely collected. An obvious example is dairy production, where several lines of data are often collected while milking the cows, but where a good standard for combining this data for meaningful information extraction is still lacking (Rutten *et al.* 2013).

## Conclusions

We show that one can meaningfully co-model three very different types of monitoring data (live weight, feed usage and water flow) from an animal production herd, using a multivariate dynamic linear model. The errors in the forecasts produced by such a model can be unified to allow for easy monitoring of the health state of the herd using a Shewhart control chart to raise appropriate alarms.

## Acknowledgements

This research was carried out with support from The Danish Council for Strategic Research (The PigIT Project, Grant number 11-116191)

## References

Bono, C., Cornou, C. & Kristensen, A.R., 2012. Dynamic production monitoring in pig herds I: Modeling and monitoring litter size at herd and sow level. *Livestock Science*, 149(3), pp.289–300.

Hogeveen, H., Kamphuis, C., Steeneveld, W. & Mollenhorst, H., 2010. Sensors and clinical mastitis--the quest for the perfect alert. *Sensors (Basel, Switzerland)*, 10(9), pp.7991–8009.

Montgomery, D., 2005. *Introduction to Statistical Quality Control* 5th ed., Hoboken, NJ, USA: Wiley.

Ostersen, T., Cornou, C. & Kristensen, A.R., 2010. Detecting oestrus by monitoring sows' visits to a boar. *Computers and Electronics in Agriculture*, 74(1), pp.51–58.

Rutten, C.J., Velthuis, A.G.J., Steeneveld, W. & Hogeveen, H., 2013. Invited review: sensors to support health management on dairy farms. *Journal of dairy science*, 96(4), pp.1928–52.

The R Core Team, 2013. *R : A Language and Environment for Statistical Computing* 2.15.3 ed., R Foundation for Statistical Computing.

Viguier, C., Arora, S., Gilmartin, N., Welbeck, K. & O'Kennedy, R., 2009. Mastitis detection: current trends and future perspectives. *Trends in biotechnology*, 27(8), pp.486–93.

West, M., Harrison, J., 1997. *Bayesian Forecasting and Dynamic Models* 2nd ed., New York, USA: Springer.

# Chapter 8: Paper 4

Manuscript

## A multivariate dynamic linear model for early warnings of diarrhea and pen fouling in slaughter pigs

Dan Børge Jensen[1,*], Nils Toft[2], Anders Ringgaard Kristensen[1]

[1]*Department of Large Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 2, DK-1870 Frederiksberg C, Denmark*
[2]*National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark*

[*]Corresponding author: daj@sund.ku.dk

### ABSTRACT
We present a method for providing early, but indiscriminant, predictions of diarrhea and pen fouling in grower/finisher pigs. We collected data on live weight, dispensed feed amount, water flow, drinking bouts frequency, temperature at two positions per pen and section level humidity from 16 pens (8 double pens) over three full growth periods. The separate data series were co-modeled at pen level with time steps of one hour, using a multivariate dynamic linear model. The step-wise forecast errors of the model were unified using Cholesky decomposition. An alarm was raised, if the unified error exceeded a set threshold a sufficient number of times, consecutively. Using this method with a seven day prediction window, we achieved an area under the receiver operating characteristics curve of 0.88, and a specificity of 0.81 when the sensitivity was 0.80. Shorter prediction windows yielded lower performances, but longer prediction windows did not affect the performance.

**Key words:** early warning, DLM, modeling, pigs, prediction

### 1 INTRODUCTION
Although it is not yet widely used in pig production, many different types of data could be collected in a standard pig herd and used in systems for early detection of disease or undesired behaviors.

For example, we have previously shown that continuously monitoring pen level temperature will yield information, which is useful for early detection of diarrhea and pen fouling (Jensen & Kristensen 2016).This makes sense, given that temperature is well known as a key factor for the onset of pen fouling (Aarnink et al. 2006), and that in general, pigs are more sensitive to the surrounding temperatures than for example cattle (Young 1981). Diurnal temperature differences in particular can cause the pigs to show symptoms of stress, as well slower growth rate and higher feed consumption (Lopez et al. 1991).

Similarly, pigs are generally known to have stable diurnal drinking patterns, from which they don't typically deviate unless they are affected by disease outbreaks or environmental stressors. A model which accurately describes these drinking patterns have been presented (Madsen et al. 2005), and a test of this model on 12 batches of growing pigs suggest that deviations from predicted water consumption can provide warning of diarrhea approximately 24 hours before it was otherwise detected (Madsen & Kristensen 2005).

Feed consumption and live weight are typically monitored closely in breeding stations, where selecting the most efficient growers is important. Here, feed will often be dispensed *ad libitum* to individual pigs, identified by RFID tags. This type of data has further been used to investigate such questions as how some environmental factors affect growth rate and efficiency (Jensen et al. 2014) and whether the rate of weight gain affects lean meat production in growing pigs (Stege et al. 2011). However, so far as we can tell, continuous weight and feed consumption data have never been tested as parameters in systems for early warnings of disease or undesired behavior. This may be because neither *ad libitum* feeding nor accurate, continuous monitoring of pig weights are common practices on most slaughter pig producing farms, and thus the utility of a system relying on this data would be limited.

We have previously shown (Jensen et al. 2015) that distinct types of data such as live weight, dispensed feed amount and water consumption could be meaningfully co-modeled using a multivariate dynamic linear model (DLM) (West and Harrison, 1997) with time steps of one day. To our knowledge, that was the first demonstration of combining such different data series with a single DLM. A few examples exists of univariate DLM's for detecting events in animal production, *e.g.* the water consumption model by Madsen et al. (2005) and for automated estrus detection in sows (Ostersen et al. 2010). Here, we intend to expand on our previously described multivariate DLM with the aim of predicting undesired events in a Danish pig herd.

The multivariate DLM produces one-step-ahead forecasts of the modeled variables, which can be compared with actual observations. When drastic changes occur in the modeled system, such as in the case of a disease outbreak, the absolute values of forecast errors will increase. The DLM is however also adaptive by design, and thus a model starting with some general assumptions can adapt to the peculiarities of a specific system, such as a specific batch of growing pigs. Furthermore, the co-dependencies between several variables of interest can be taken into account, when one-step-ahead forecasts for these variables are calculated.

We hypothesize that when monitoring a group of pigs with a model which is optimized to describe the pigs under normal and healthy conditions, the model will be able to accurately predict new observations, so long as the pigs remain healthy. Therefore, when the model is unable to provide accurate forecasts, the pigs have either changed or are in the process of changing to an abnormal state. We unify the individual forecast errors produced at each observation using Cholesky decomposition. We therefore call this method the *DLM/Cholesky method*.

Here, we intend to show that the DLM/Cholesky method, including live weight, feed amount, water consumption, drinking bouts frequency, pen level temperature, and section level humidity in the DLM, can be used to accurately, but indiscriminately, predict diarrhea and pen fouling at pen level. We further intend to estimate the relative information value of these observable variables with respect to such early warnings.

# 2 MATERIALS AND METHODS

## 2.1 Data source

All data used in this study were obtained from the finisher unit of a commercial Danish pig farm, housing slaughter pigs as they grow from approximately 30 to 100 kg. The unit consists of five sections, each with 14 pens. Each pen contains 18 pigs (at insertion), sorted by sex and size. The climate within each section is controlled by a combi-diffuse ventilation system, computer-controlled sprinklers above each pen and heating pipes installed in the back walls. All pens in the section are paired into double pens, where two neighboring pens share feed and water supplies.

Sensors were installed to automatically record data on feed usage, water flow to the drinking nipples and temperature in 16 of the 70 pens in the finisher unit, and humidity was monitored on section level. Furthermore, the pigs from two double pens (four pens) from one section (section 2) of the unit were weighed manually once per week.

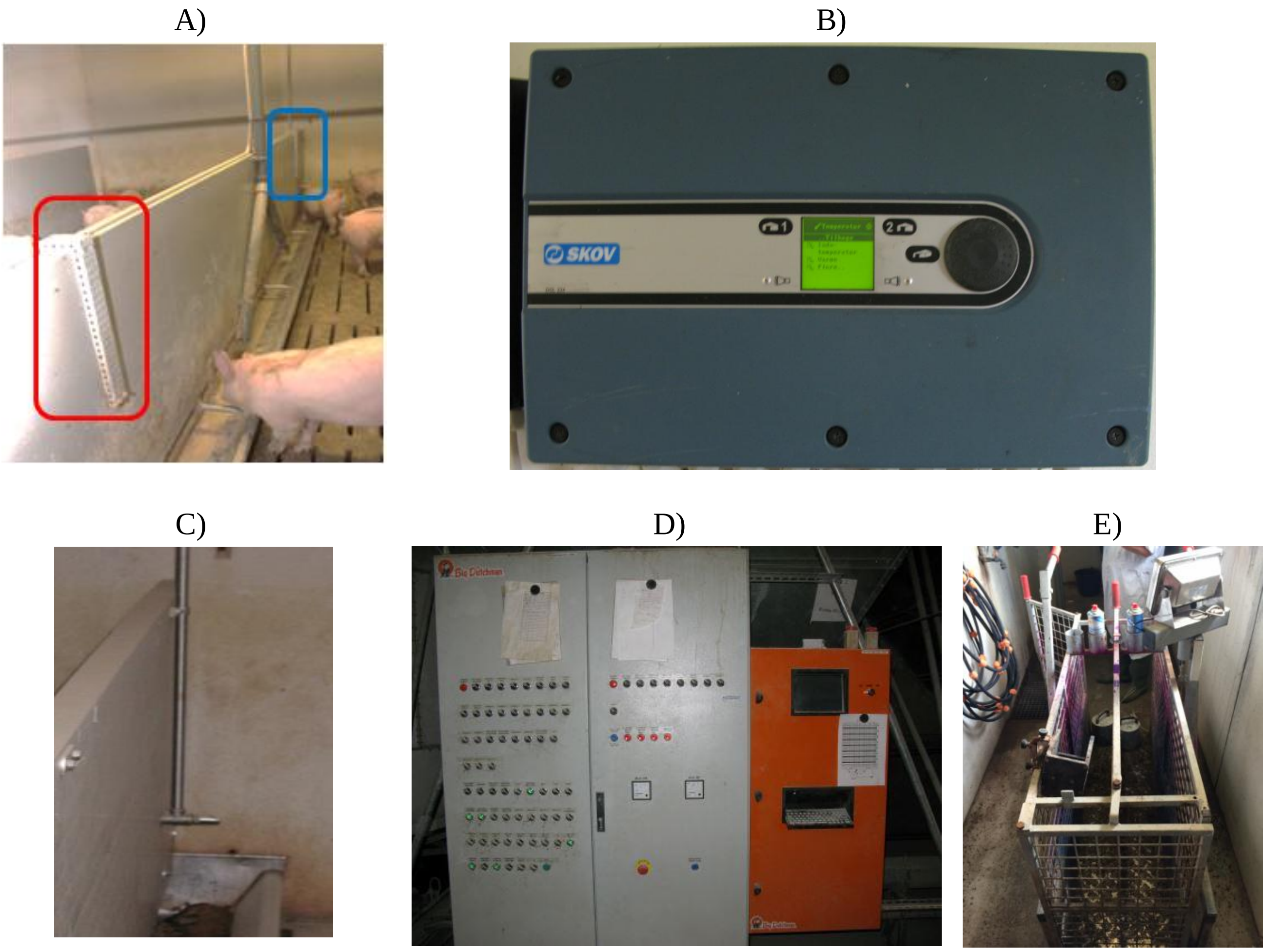



**Figure 1: The sensors used to monitor the growing pigs. A) Two thermometers are placed in each pen, one at the back wall (blue rectangle) and one at the corridor (red rectangle). Temperatures are measured at single pen level. B) The climate computer, from which we got data on section level humidity. C) Water is dispensed from a drinking nipple. A single pipe supplies both pens in a double pen. The water flow is measured at double pen level. D) The feeding system, responsible for dispensing predetermined amounts of wet feed to a double pen per day. E) Scale for weekly manual weighing of individual pigs from section two.**

Temperatures were measured continuously at the single pen level. Two thermometers were installed in each pen, one near the back wall of the pen, one near the section corridor (Figure 1 A). The temperatures were measured at two positions in every pen: near the back wall, *i.e.* near the designated resting area with solid floors, and near the section corridor, *i.e.* near the designated excretion area with slatted floors. The temperature near the back wall was generally a few degrees higher than near the corridor.

Humidity data were retrieved from the climate computer (B). Humidity data was generally available once per day. However, during one spring time growth period, continuous humidity recordings were available.

Water flow was measured continuously at the double pen level. Each pen in a double pen has a drinking nipple, but both share a single water pipe (Figure 1 C). A flow meter was installed in each of these pipes, where it measured rotations per second. A separate calibration was done for each water pipe.

Feed was dispensed per double pen by a feeding system from the company Big Dutchman (Figure 1 D). Feed was dispensed daily according to a preset feeding curve. However, when relevant, the amounts would be adjusted daily based on the observed surplus of feed not eaten on a given day.

The individual pigs in section 2 were weighed manually using the scale depicted in Figure 1 E. The weightings were performed from the week of insertion until the first pigs from that section were sent to the abattoir. The pigs were individually identified with RFID ear tags, and it was known to which single pen an individual pig belonged.

Lastly, observations of undesired events were observed at single pen level on a daily basis. This was done manually by the farm staff, who noted any undesired event in a log book as part of their daily routine. This study focuses on two specific undesired events, namely diarrhea and pen fouling, and hence forth "events" shall refer specifically to these two types of observations.

Data was collected between November 20$^{th}$ 2013 and December 12$^{th}$ 2014, during which time three new batches of pigs were inserted in each pen.

## 2.2 Data editing

All data editing, modeling, and various calculations were done using the statistical language and environment R (The R Core Team, 2013).

The collected data were scanned for outliers using simple summary statistics. Humidity and temperature were found to be plagued by values below 1 percent and centigrade, respectively. In addition, humidity was plagued with values above 100 %. Such values were considered as missing observations during the modeling described below. All other variables were found to be within acceptable ranges. The water flow and drinking bouts frequency data were particularly plagued by long periods of missing observations. From previous work (Madsen & Kristensen 2005) we had reason to think that water flow would be particularly important for the detection of the undesired events. For this reason, we only included batches with less than 60 % missing water flow observations when testing our model.

The continuously measured variables, *i.e.* water flow, temperature, and humidity when relevant, were aggregated to one-hour mean values. From the water flow data, we further calculated drinking bouts frequency, *i.e.* how many times the water nipple was activated per pig per (double) pen during any given hour. This value served as a proxy for pig activity level.

The variables measured only once per day or less, *i.e.* live weight, feed amount, and humidity when relevant, had no specific time stamp associated with them. In all such cases it was assumed that these observations were made at noon.

The event observations, also without timestamps, were always assumed to be observed at the zero'th our (midnight) of the day where they were observed.

Table 1 shows the descriptive statistics of the data used in this study, following the above mentioned data editing.

**Table 1: Descriptive statistics of the data, used in the present study**

| | Outcome variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | No event | | Any event | | Diarrhea | | Pen fouling | |
| N | 4146 | | 36 | | 17 | | 19 | |
| Predictive variable | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| LW[1] | 51.9 | 17.4 | 66.8 | 13.5 | 57.4 | 22.7 | 71.6 | 6.6 |
| Feed amount[2] | 15.6 | 6.3 | 16.4 | 7.8 | 14.7 | 5.2 | 17.9 | 9.4 |
| Water flow[3] (day[4]) | 0.2 | 0.2 | 0.2 | 0.3 | 0.2 | 0.2 | 0.2 | 0.3 |
| Water flow[3] (night[5]) | 0.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 |
| Drinking bouts frequency[6] (day[4]) | 12.6 | 16.1 | 13.7 | 17.8 | 14.8 | 18.0 | 12.6 | 17.7 |
| Drinking bouts frequency[6] (night[5]) | 2.9 | 5.6 | 3.2 | 4.8 | 3.9 | 5.6 | 2.6 | 3.9 |
| Temperature[7], corridor (day[4]) | 20.1 | 2.9 | 20.1 | 2.9 | 19.7 | 2.8 | 20.3 | 2.2 |
| Temperature[7], corridor (night[5]) | 19.3 | 2.3 | 19.5 | 2.1 | 19.1 | 2.3 | 19.8 | 1.8 |
| Temperature[7], back wall (day[4]) | 21.7 | 2.7 | 21.6 | 2.7 | 21.7 | 2.8 | 21.6 | 2.6 |
| Temperature[7], back wall (night[5]) | 21.7 | 2.7 | 21.4 | 1.9 | 21.7 | 1.7 | 21.2 | 2.1 |
| Humidity[8] | 73.1 | 6.3 | 84.4 | 14.9 | 81.0 | 14.1 | 86.9 | 15.4 |

[1]Average live weight in kg per pig per pen, over the entire growth period
[2]Average feed amount in kg dispensed per pig per pen, over the entire growth period
[3]Liter per hour per pig per pen
[4]The hours from and including 10 AM to 8 PM
[5]The hours from and including 9 PM to 9 AM
[6]Number of water nipple activation per hour per pig per pen
[7] °C
[8] %

Lastly, the data were split into a learning set and a test set. The learning set was used to estimate the variance components and the initial mean vector for the DLM, as described below, and only included those batches where no events were observed (N=26). The test set was used to evaluate the predictive performance of the DLM/Cholesky method presented in this study. It consisted only of those batches were an undesired event was observed at least once and where less than 60 % of water observations were missing. Thus 6 batches with events were omitted while 16 were included. The included batches contained a total of 12 diarrhea events and 13 pen fouling events.

## 2.3 Application of dynamic linear model

A multivariate DLM with one step Markov evolution (West and Harrison, 1997) were used to co-model the seven observed variables (live weight, feed amount, corridor temperature, back wall temperature, water flow, drinking bouts frequency, and humidity). In general, a DLM consists of an observation equation and a system equation (Equations 1 and 2, respectively) as follows:

$$\boldsymbol{Y_t} = \boldsymbol{F_t'}\boldsymbol{\theta_t} + \boldsymbol{v_t}, \qquad \boldsymbol{v_t} \sim N(\underline{0}, \boldsymbol{V}) \tag{1}$$

$$\boldsymbol{\theta_t} = \boldsymbol{G_t}\boldsymbol{\theta_{t-1}} + \boldsymbol{w_t}, \qquad \boldsymbol{w_t} \sim N(\underline{0}, \boldsymbol{W}) \tag{2}$$

Equation 1 describes how the values of an observation vector $(\boldsymbol{Y_t})$ depend on an unobservable parameter vector $(\boldsymbol{\theta_t})$ at time $t$. The design matrix $(\boldsymbol{F_t'})$ serves to separate the true values of the observed variables from their inferred rate of change, *i.e.* their trends. $\boldsymbol{V}$ is the observational co-variance matrix, describing the co-variance between the observed variables.
Equation 2 describes how the parameter vector $(\boldsymbol{\theta_t})$ is updated from time $t-1$ to time $t$. The system matrix $(\boldsymbol{G_t})$ serves to facilitate this updating by adding the inferred trends to the estimated true value of the observable variables. $\boldsymbol{W}$ is the systematic co-variance matrix, describing the co-variance between the evolution of the observed variables as well as their trends.

To help decide how the individual sensor variables should be handled by the model, plots based on the healthy batches in the learning set data were made to display the average values of each variable given an appropriate time frame, as shown below. When referencing seasons, winter is defined here as December-January-February, spring is Marts-April-May, summer is June-July-August and autumn is September-October-November. On all plots, the color codes are: Blue = winter, green = spring, red = summer, black = autumn.

To ensure that the variances of the seven different variables were approximately at the same level when co-modeling, the values of live weight, feed usage, drinking bouts frequency, and humidity were divided by 10, water flow was multiplied by 10, and the two temperature variables were kept unchanged.

### 2.3.1 Live weight

The average live weights given the four seasons are seen on Figure 2.

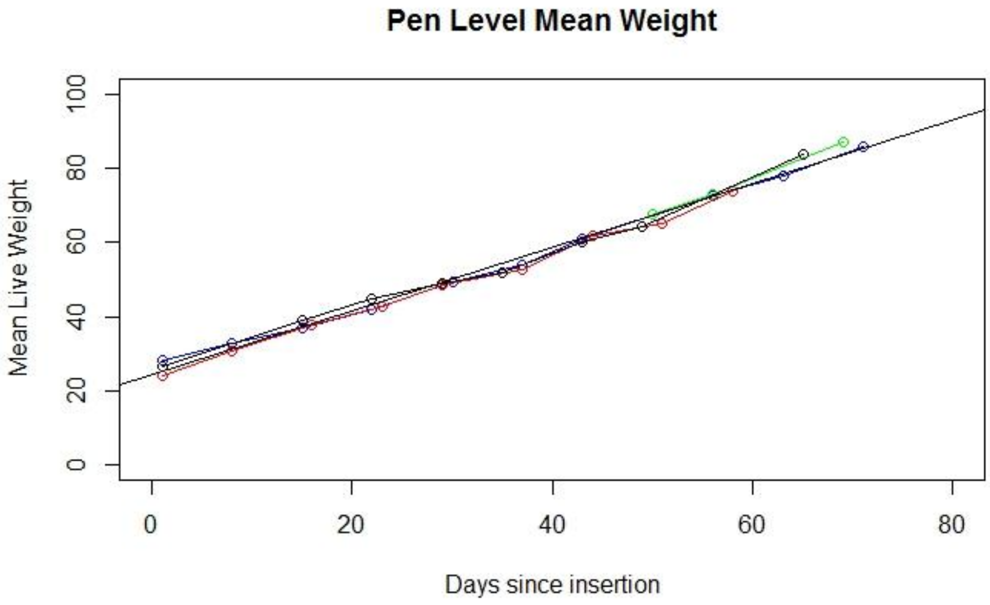



**Figure 2: The mean live weight over a growing period given the four seasons of the year. The mean live weight in any pen, regardless of season, is initially assumed to follow the linear model represented by the straight, black line is. Blue = winter, green = spring, red = summer, black = autumn.**

As is seen, the mean weights approximately follow a linear function with no apparent difference in the growth between the different seasons of the year. Thus the mean weights were always assumed to initially follow the same linear function with an initial mean weight of 26.2 kg and a growth rate of 6 kg/week, approximately 0.04 kg/hour, illustrated as the straight line on Figure 2.

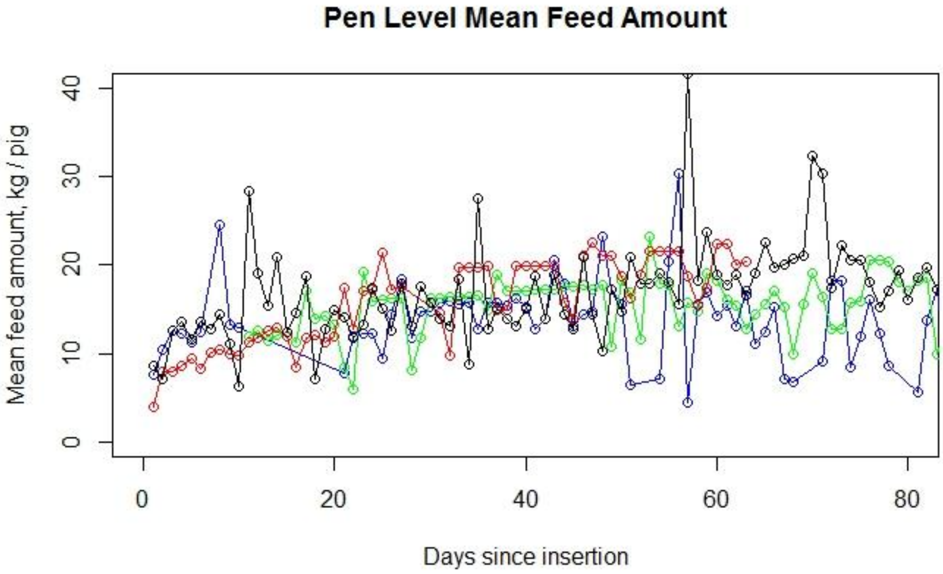
Thus live weight alone could be modeled using the initial parameter vector ($\boldsymbol{\theta_0}$), the design matrix ($\boldsymbol{F'_t}$), and the system matrix $\boldsymbol{G_t}$ seen in Table 2.

**Table 2: The initial parameters needed for a dynamic linear model of live weight alone**

| $\boldsymbol{\theta_0}$ | $\boldsymbol{F'_t}$ | $\boldsymbol{G_t}$ |
|---|---|---|
| $\begin{bmatrix} 2.62 \\ 0.004 \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 \end{bmatrix}$ | $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ |

### 2.3.2 Feed amount

The mean feed amount dispensed per pig in a double pen over a growth period, given the training set, is seen on Figure 3




.

**Figure 3: Mean feed amount per day over the growing period, given the four season of the year. Blue = winter, green = spring, red = summer, black = autumn.**

As is seen, there is much variation in the amount of feed dispensed per day, even when averaging over all included batches. Nevertheless, the initial amount is always relatively low (8.1 kg/pig), while the final amount is relatively high (~20 kg/pig), and no consistent differences are seen between the four seasons. Thus it was decided to model the feed amount as a linear function of time anfter insertion with an initial level of 8.1 kg/pig and an initial trend of 0.06 kg/hour (~9.3 kg/week). Thus the DLM parameters for feed alone would be as seen in Table 3.

**Table 3: The initial parameters needed for a dynamic linear model of dispensed feed amount alone**

| $\boldsymbol{\theta_0}$ | $\boldsymbol{F'_t}$ | $\boldsymbol{G_t}$ |
|---|---|---|
| $\begin{bmatrix} 0.81 \\ 0.006 \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 \end{bmatrix}$ | $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ |

### 2.3.3 Temperature

Figure 4 shows the average temperatures per hour between midnight and midnight, given the four different seasons. It is clearly seen that temperatures near the back wall are generally a few degrees higher than near the corridor.

A) B)

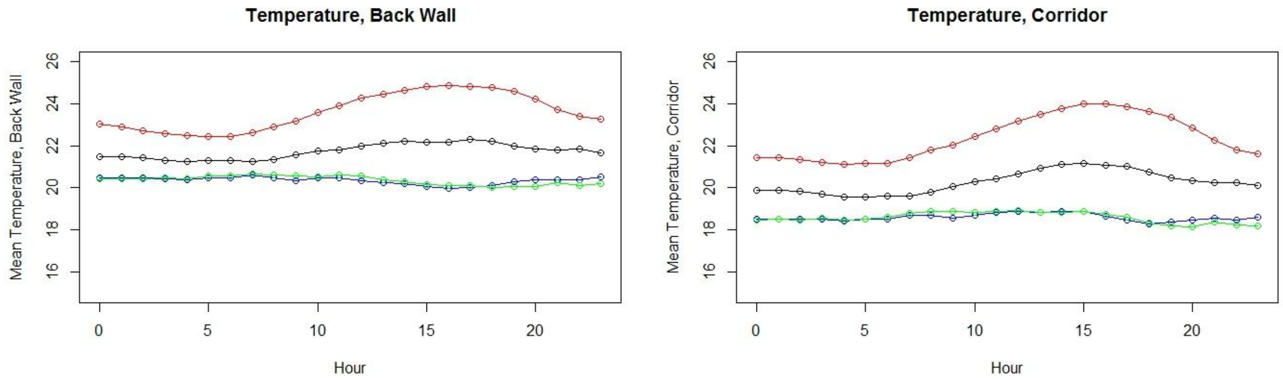



**Figure 4: The average temperatures between midnight and midnight, given the season of the year, near the back wall (A) and near the corridor (B). Blue = winter, green = spring, red = summer, black = autumn.**

At both positions, the average temperatures for each season can be described by a season-specific harmonic wave. The system matrix describing a harmonic wave is seen in Equation (3)

$$\boldsymbol{G} = \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{bmatrix} \quad (3)$$

, where $\omega$ is the period. In our case, the waves have a period of 24 hours, meaning that $\omega = (2\pi)/24$.

Due to the clear differences between the four seasons, a separate initial mean vector, $\boldsymbol{\theta_0}$, was estimated for each of the seasons. This was done by having a single common mean vector of $\boldsymbol{\theta_0^{Position}} = \left[T_0^{Position}, 1, 1\right]'$, where $T_0^{Position}$ is the overall average temperature at midnight for the relevant position, *i.e.* by the corridor or by the back wall. Using this initial mean vector for each season, each of the 14 healthy batches in the learning set was modeled with a DLM and subsequently smoothed, as described by West and Harrison (1997). During the modeling, the systematic variance, $\boldsymbol{W}$, of the temperature was not yet known, and it was thus assumed that all variance was accounted for by the observational variance, $\boldsymbol{V}$. This variance was calculated directly from the temperature observations of the healthy batches in the learning set, using the R function var().

As a result of the modeling and subsequent smoothening, a new estimate of the initial mean vector given the season was produced with the modeling and smoothening of each healthy batch. For each healthy batch, the new initial mean vector was saved and used when modeling the next batch, for which observations started during the same season. Thus the $\boldsymbol{\theta_0^{Season}}$ vectors were iteratively improved. Their final form can be seen in Table 4. Notice that the two $\boldsymbol{\theta_0^{Spring}}$ vectors retained their original form, as no batches, for which data were available, had their initial observations during the spring.

**Table 4: The initial mean vectors given season and sensor placement within the pen.**

| | $\boldsymbol{\theta}_0^{Winter}$ | $\boldsymbol{\theta}_0^{Spring}$ | $\boldsymbol{\theta}_0^{Summer}$ | $\boldsymbol{\theta}_0^{Autumn}$ |
|---|---|---|---|---|
| Corridor | $\begin{bmatrix}17.1\\0.003\\0.087\end{bmatrix}$ | $\begin{bmatrix}19.0\\1\\1\end{bmatrix}$ | $\begin{bmatrix}18.7\\-0.287\\0.221\end{bmatrix}$ | $\begin{bmatrix}15.7\\0.024\\0.035\end{bmatrix}$ |
| Back wall | $\begin{bmatrix}19.3\\0.002\\-0.015\end{bmatrix}$ | $\begin{bmatrix}21.3\\1\\1\end{bmatrix}$ | $\begin{bmatrix}21.3\\-0.192\\0.146\end{bmatrix}$ | $\begin{bmatrix}18.0\\0.033\\-0.065\end{bmatrix}$ |

The $\boldsymbol{G}$ and $\boldsymbol{F_t}$ matrices, which would be used when describing the temperatures at both positions, are seen in

**Table 5: The design matrix and system matrix used for describing the temperatures in the pen, regardless of sensor position.**

| $\boldsymbol{F}_t'$ | $\boldsymbol{G}$ |
|---|---|
| $[1 \quad 0 \quad 0]$ | $\begin{bmatrix}1 & 1 & 0\\0 & \cos(\omega) & \sin(\omega)\\0 & -\sin(\omega) & \cos(\omega)\end{bmatrix}$ |

### 2.3.4 Water flow and drinking bouts frequency

When estimating the initial parameters for the drinking behavior (*i.e.* water flow and drinking bouts frequency), only batches with at less than 60 % missing observations of the drinking behavior data were included (24 of the 26 healthy batches).

Figure 5A shows the mean water flow in liters per pig from midnight to midnight, while Figure 5B shows the mean frequency of water nipple activations between midnight and midnight.

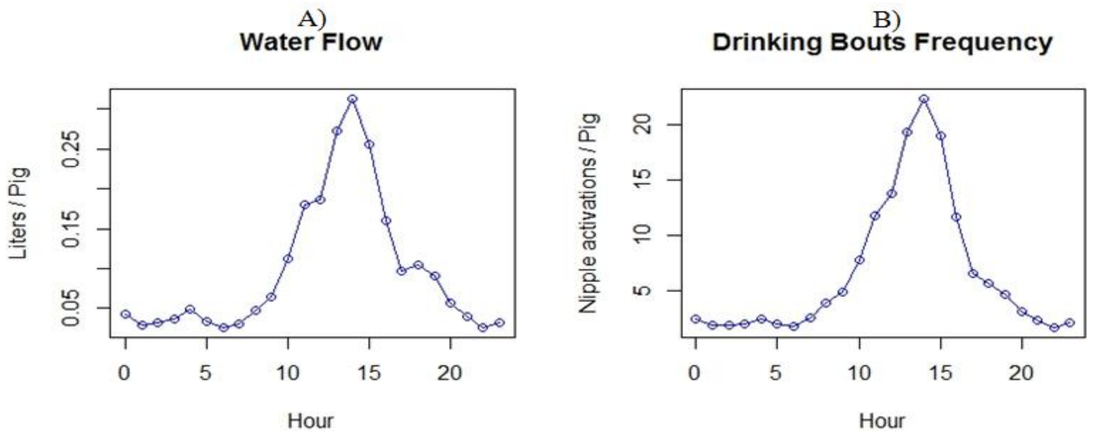



**Figure 5: A) average water flow (L/Pig) per hour between midnight and midnight. B) The drinking bouts frequency, *i.e.* the number of water nipple activations per pig, between midnight and midnight**

The daily drinking pattern observed in our data (Figure 5 A) is similar to the pattern modeled by Madsen et al. (2005), in spite of the two data sets being completely independent. Madsen et al. (2005) showed that their data were best described as the sum of three harmonic waves, and it was decided to follow their example. Thus the design matrix and the system matrix, which were used to model both water flow and drinking bouts frequency, were as is seen in Table 6.

**Table 6: The design matrix and system matrix used for describing the water flow and drinking bouts frequency.**

| $\boldsymbol{F}_t'$ | $\boldsymbol{G}$ |
|---|---|
| $\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}$ | $\begin{bmatrix}1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos(\omega) & \sin(\omega) & 0 & 0 & 0 & 0 \\ 0 & 0 & -\sin(\omega) & \cos(\omega) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos(2\omega) & \sin(2\omega) & 0 & 0 \\ 0 & 0 & 0 & 0 & -\sin(2\omega) & \cos(2\omega) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cos(3\omega) & \sin(3\omega) \\ 0 & 0 & 0 & 0 & 0 & 0 & -\sin(3\omega) & \cos(3\omega)\end{bmatrix}$ |

As is seen from the system matrix, $\boldsymbol{G}$, in Table 6, the water flow and drinking bouts frequency has an overall trend combined with three harmonic waves.

As with the temperature models, the initial mean vector was estimated by first modeling the healthy water flow and drinking bouts frequency observations from the learning set while assuming that all variance was accounted for by the observational variance, $\boldsymbol{V}$, and subsequently smoothening the filtered observations. The resulting initial mean vectors are seen in Table 7.

**Table 7: Initial mean vectors of water flow and drinking bouts frequency.**

| | Water flow | Drinking bouts frequency |
|---|---|---|
| $\boldsymbol{\theta}_0$ | $\begin{bmatrix}0.207 \\ 0.000 \\ -0.105 \\ 0.328 \\ 0.099 \\ -0.321 \\ -0.052 \\ 0.117\end{bmatrix}$ | $\begin{bmatrix}0.066 \\ 0.000 \\ -0.055 \\ 0.236 \\ 0.111 \\ -0.213 \\ -0.066 \\ 0.096\end{bmatrix}$ |

#### 2.3.5 Humidity

The average section level humidity given the hour of the day can be seen on Figure 6.

. Notice that hourly humidity observations only exist in our data set for one spring time growing period. For all other growing periods, humidity is generally only available once per day. These observations are assumed to be made at noon.

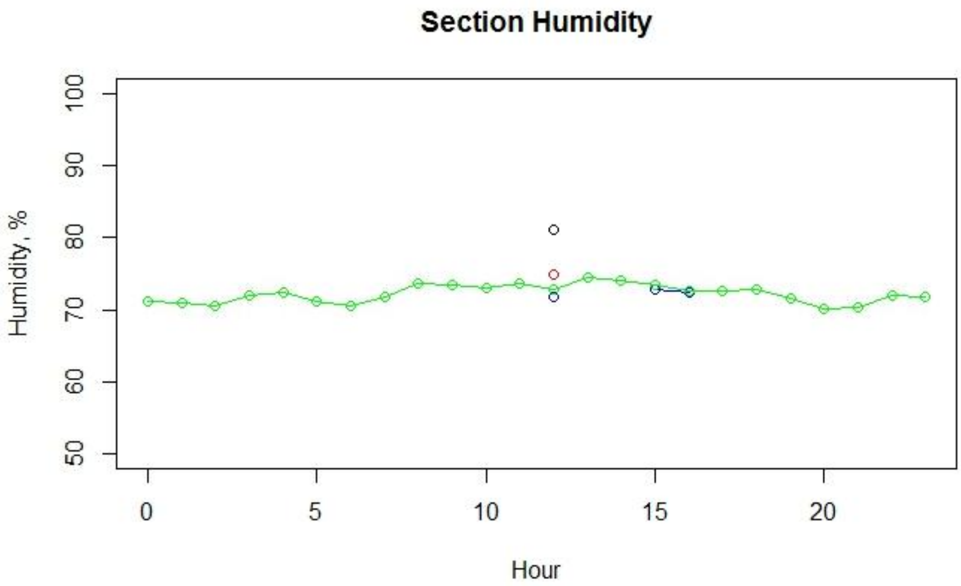



**Figure 6: The mean humidity observation given hour between midnight and midnight. Hourly humidity observations are only available for a single growing period during the spring time. All other humidity observations are made once per day, which is assumed to be at noon. Blue = winter, green = spring, red = summer, black = autumn.**

As is seen, the observed hourly humidity values are relatively constant at between 70 and 75 % between midnight and midnight. Although the mean autumn humidity (~80 %) seems higher than the mean humidity values of the other seasons (~70-75 %), it was decided not to include separate humidity sub-models for the four different seasons. Instead it was decided to model humidity as a single linear function with a starting value estimated as the mean value of all humidity observations said to be observed at noon after data editing (mean humidity = 75.4 %) with a trend of 0. Thus the initial DLM parameters for humidity alone were as seen in Table 8.

**Table 8: Initial parameters for DLM of humidity alone.**

| $\boldsymbol{\theta_0}$ | $\boldsymbol{F_t'}$ | $\boldsymbol{G_t}$ |
|---|---|---|
| $\begin{bmatrix} 7.54 \\ 0 \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 \end{bmatrix}$ | $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ |

The trend of 0 means that we expect the humidity not to change, but that we allow for the possibility that a trend can be learned during the run of the DLM.

#### 2.3.6 Combined model

To co-model the seven different observable variables, the DLM parameters corresponding to the separate variables, described above, were combined into a single model. Thus the combined initial parameter vector, $\boldsymbol{\theta_0}$, was a column vector with a length of 28. Because the initial mean values related to temperature depended on the season of the year, the appropriate $\boldsymbol{\theta_0}$ vector was defined before modeling each batch.

The design matrix, $\boldsymbol{F_t'}$, only contains the rows corresponding to the variables observed at a given time, $t$. Thus when all variables are observed at the same time, $\boldsymbol{F_t'}$ has a 7x28 structure, while when only one variable is observed, $\boldsymbol{F_t'}$ has a 1x28 structure.

$\boldsymbol{G}$ had a constant 28x28 structure.

## 2.4 Estimation of variance matrices

The observational and systematic co-variance matrices ($\boldsymbol{V}$ and $\boldsymbol{W}$, respectively) were estimated using the expectation maximization (EM) algorithm (West and Harrison, 1997), on the healthy batches in the learning set. By plotting the diagonal values of $\boldsymbol{V}$ and $\boldsymbol{W}$ against the number of iterations, visual inspection was used to determine when the algorithm had converged.

## 2.5 Unification of forecast errors

In this study, the forecast errors produced by the DLM were unified in the same way as was done in (Jensen et al. 2015). Specifically, during the modeling, a vector $(e_t)$ of forecast errors (*Observed values – Forecasted values*) was continuously generated for each time step. In addition, a matrix describing the forecast co-variances $(Q_t)$ was continuously generated, as described by West and Harrison (1997). The Cholesky decomposition was calculated for $Q_t$ using the R function chol. Using the decomposed matrix $\left(C_{Qt}\right)$, the error vector was transformed, as seen in Equation (4).

$$u_t = C_{Qt}^{-1} \cdot e_t \tag{4}$$

This transformation ensures that the transformed error values in $u_t$ are mutually independent and each follow a standard normal distribution. Thus a single value measuring the square of the deviation from 0 can be easily calculated for the set of forecast errors, as seen in Equation (5).

$$d_t^2 = u_t' \cdot u_t \tag{5}$$

This unified error will follow a $\mathcal{X}^2$ distribution with $n$ degrees of freedom, where $n$ is the number of elements in $u_t$, *i.e.* the number of variables observed at time $t$. Thus $d_t^2$ can be plotted to a conventional Shewhart control chart (Montgomery 2005) to allow for an easy monitoring of the complex system. Several quintiles for the $\mathcal{X}^2$ distribution were tested as the basis of alarm control limits, as described in detail below.

To allow for a constant control limit in response to varying degrees of freedom, $d_t^2$ was adjusted, according to Equation (6).

$$d^2_{adj.t} = d^2_t \cdot \left(\frac{\mathcal{X}^2(Quantile, 7)}{\mathcal{X}^2(Quantile, n)}\right) \tag{6}$$

A notable risk associated with the use of Cholesky decomposition is that the DLM will fail if $Q_t$ is (computationally) singular or not positive definite. When running the DLM on the test set, such errors were handled by setting the unified error to 1, effectively meaning that no alarm would be associated with any observation which caused this error. This problem was irrelevant when estimating the variance components using the EM algorithm.

## 2.6 Performance evaluation

Part-alarms were said to be raised if the unified error exceeded the control limit, as defined by the quintiles for the $\mathcal{X}^2$ distribution and the degrees of freedom. Full alarms were said to be raised if a sufficient number of part-alarms were raised consecutively. Predictive performance was estimated based on prediction windows around the days with observed undesired events. A review by Hogeveen et al. (2010) describes this as a standard method for evaluation the performance of mastitis detection in dairy cows, and we have previously used the method to evaluate a temperature-based model for early warning of diarrhea and pen fouling in slaughter pigs (Jensen & Kristensen 2016). A number of different windows were tested. Figure 7 illustrates the method when using one of these windows, which includes three days before and one day after an event observation, hence forth designated as a -3/+1 window. In this case, if at least one full alarm occurred at most three days before or one day after the day of an event observation, that was counted as a true positive (TP). If no full alarms were raised within such a prediction window, it was counted as a false negative (FN). Days with full alarms which occurred outside of a prediction window were counted as false positives (FP), and days without full alarms which were outside a prediction window were counted as true negatives (TN).

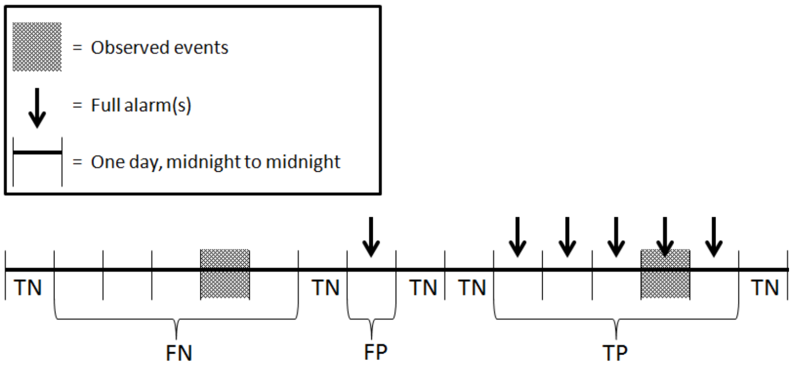



**Figure 7: Hypothetical examples to how true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are defined. In this illustration we use the -3/+1 prediction window, but other windows were tested as well. When using the -3/+1 window, any number of full alarms occurring at most three days before or one day after an event observation is counted as one TP. If no full alarms occur within this window, it is counted as one FN. Single days without events but with full alarms are counted as FP. Single days with no events and no full alarms are counted as TN.**

Having counted up the TP's, FP's, TN's, and FN's, sensitivity and specificity were calculated as $Sensitivity = TP/(TP + FN)$ and $Specificity = TN/(TN + FP)$. By varying the number of consecutive part alarms required for one full alarm between 0 and 25, a receiver operating characteristics (ROC) (Zweig & Campbell 1993) curve was made. The area under the curve (AUC) served as the value, by which the predictive performance was measured. AUC was calculated using the function auc from the library MESS in R.

## 2.7 Testing part-alarm threshold

To determine the control line, above which unified errors would produce a part-alarm, all quantile values between 0.05 and 0.95 (by steps of 0.05) as well as 0.99 were tried for the $\mathcal{X}^2$ distribution of the unified errors. A ROC curve was made for each of the 20 control lines. The control line which resulted in the ROC curve with the highest AUC was selected to be used with the test set.

## 2.8 Sensitivity analysis

To investigate the relative contributions of the various variables to the performance of the described DLM/Cholesky method, each of the variables were left out during modeling and performance evaluation. The performance was then evaluated based on the AUC of the ROC curve. Furthermore, those variables, which caused the greatest reduction in AUC by their omission, were tested alone.

# 3 RESULTS AND DISCUSSION

## 3.1 Variance components

The EM algorithm was initially allowed to run for 50 iterations, at which point the visual inspection was done. This inspection revealed that the variance values had converged essentially immediately and no further iterations were run. The systematic co-variance matrix was found to be too large (dimensions: 28x28) to meaningfully include in the pages of this paper. Instead both matrices were submitted as supplementary materials, along with a matrix showing the observational correlations between the 7 variables.

## 3.2 Part-alarm threshold

Figure 8 shows the AUC values of the ROC curves achieved on the test set, when using different quantiles for the $\mathcal{X}^2$ distribution, which determines the control line for the alarms. All values are re made with a prediction window of -3/+1 days. The highest AUC (0.83) is achieved with a quantile value of 0.70. This value has been used in the generation of all results from this point on.

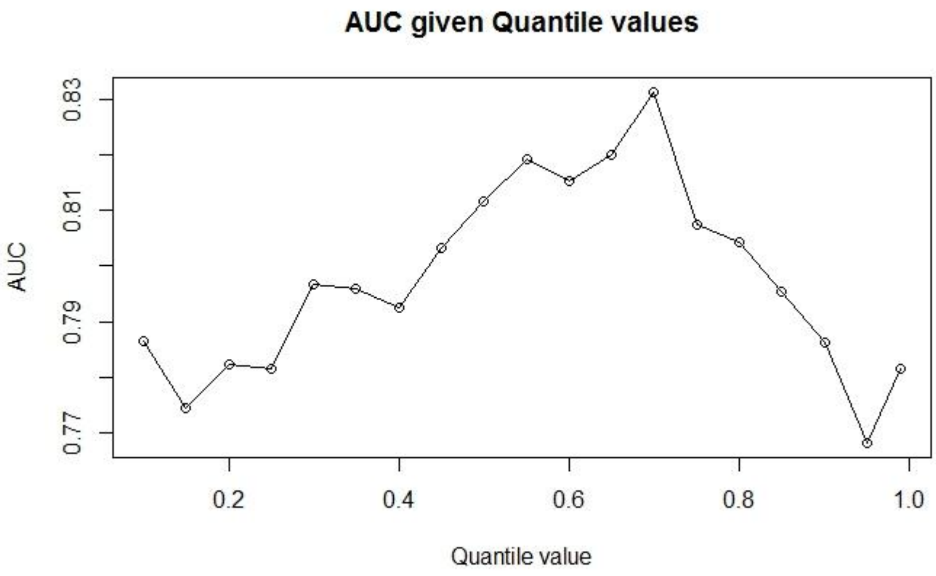



**Figure 8: AUC given quantile values for the $\mathcal{X}^2$ distribution of the unified forecast errors**

## 3.3 Model output

Figure 10 shows the unified forecast errors given the observation time for a group of pigs, for which no diarrhea or pen fouling events were observed. The red line marks the control line given the quantile of 0.70 (9.04). As is seen, it is not uncommon for the unified errors to exceed the control line, even when neither diarrhea nor pen fouling is observed. This is why multiple consecutive part-alarms are required to form one full alarm.

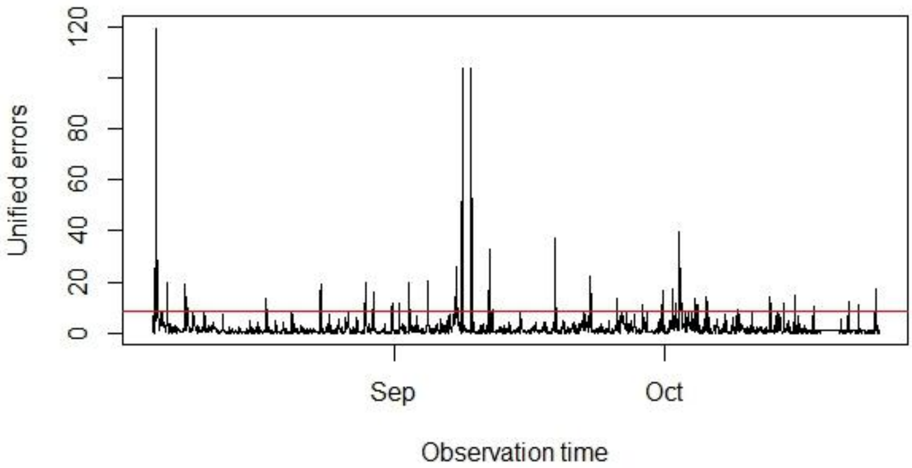



**Figure 9: Unified forecast errors for a single healthy batch of pigs.**

Table 9 shows the predictive performance of event predictions given different prediction windows. As above, *e.g.* -3/+1 means that the prediction window covers observations up to three days before and up to one day after the day on which a given event was observed. In the same way, -0/+0 means that only full alarms raised on the same day as an event was observed would count as a true positive alarm. As is seen, this one-day window results in a predictive performance which is just barely better than random guessing (AUC = 0.57), while including full alarms raised up to five days before an event observation raises the performance to 0.88. The performance is not further improved by including more than five days before the event observation. This fact suggest that true deviations from *e.g.* normal drinking behavior may start up to five days before a problem becomes apparent, or that the effect of *e.g.* sudden temperature changes does not manifest as visually identifiable problems until several days after the change. Table 9 further shows that while some information is clearly gained from including the +1 part of the window, *i.e.* allowing events to be detected up to one day after problems are physically visible, the majority of the information is actually contained in the days before the event.

**Table 9: Predictive performance given several different prediction windows. As an example, -3/+1 means that if a full alarm is raised at most three days before or one day after the day where an event was observed, that whole period is counted as one true positive prediction.**

| Prediction window | AUC |
|---|---|
| -0/+0 | 0.57 |
| -0/+1 | 0.66 |
| -1/+1 | 0.72 |
| -2/+1 | 0.81 |
| -3/+0 | 0.78 |
| -3/+1 | 0.83 |
| -4/+1 | 0.87 |
| -5/+0 | 0.84 |
| -5/+1 | 0.88 |
| -6/+1 | 0.88 |

Figure 10A shows five ROC curves corresponding to selected AUC values in Table 9, including the one produced with the -5/+1 window (triangles). This curve shows that if the sensitivity is held 0.80, as recommended for *e.g.* detection of diseases in dairy cattle by Hogeveen et al. (2010), the false positive rate would be 0.19, corresponding to a specificity of 0.81.

Table 10 shows the effect on the predictive performance of the DLM/Cholesky method when each of the variables are left out of the model. All performances in Table 10 are achieved with a prediction window of -5/+1 days, and the control line being defined by the 0.70 quantile of the $\mathcal{X}^2$ distribution for the unified forecast errors. ROC curves corresponding to selected AUC values are seen in Figure 10B.

As is seen, omitting mean live weight, feed usage, and humidity all have no discernible effect on the predictive performance, as the AUCs remain 0.88 in these cases. This matches Figure 10B, where the ROC curves given the omission of each of these variables are essentially indistinguishable from the ROC curve achieved when all variables are included.

Omitting only the temperature measurements by the corridor apparently does not influence the performance either, whereas omitting the temperature data collected by the drinking nipple reduces the AUC by three percentage points. Omitting both temperature measures simultaneously has the same effect. This is somewhat consistent with some of our previous findings (Jensen & Kristensen 2016), where five temperature summary variables were used to model the risk of diarrhea or pen fouling with a logistic regression model. Three of the five summary variables pertained to the temperature measurements by the drinking nipple, and two pertained to measurements by the corridor. Of these two variables, only one was statistically significant ($p < 0.05$).

Both drinking behavior measures, *i.e.* water flow and drinking bouts frequency, are seen to affect the performance when omitted, reducing the AUC by 3 and 1 percentage points, respectively. When both of these variables are omitted, however, a more drastic decrease of 15 percentage points is seen.

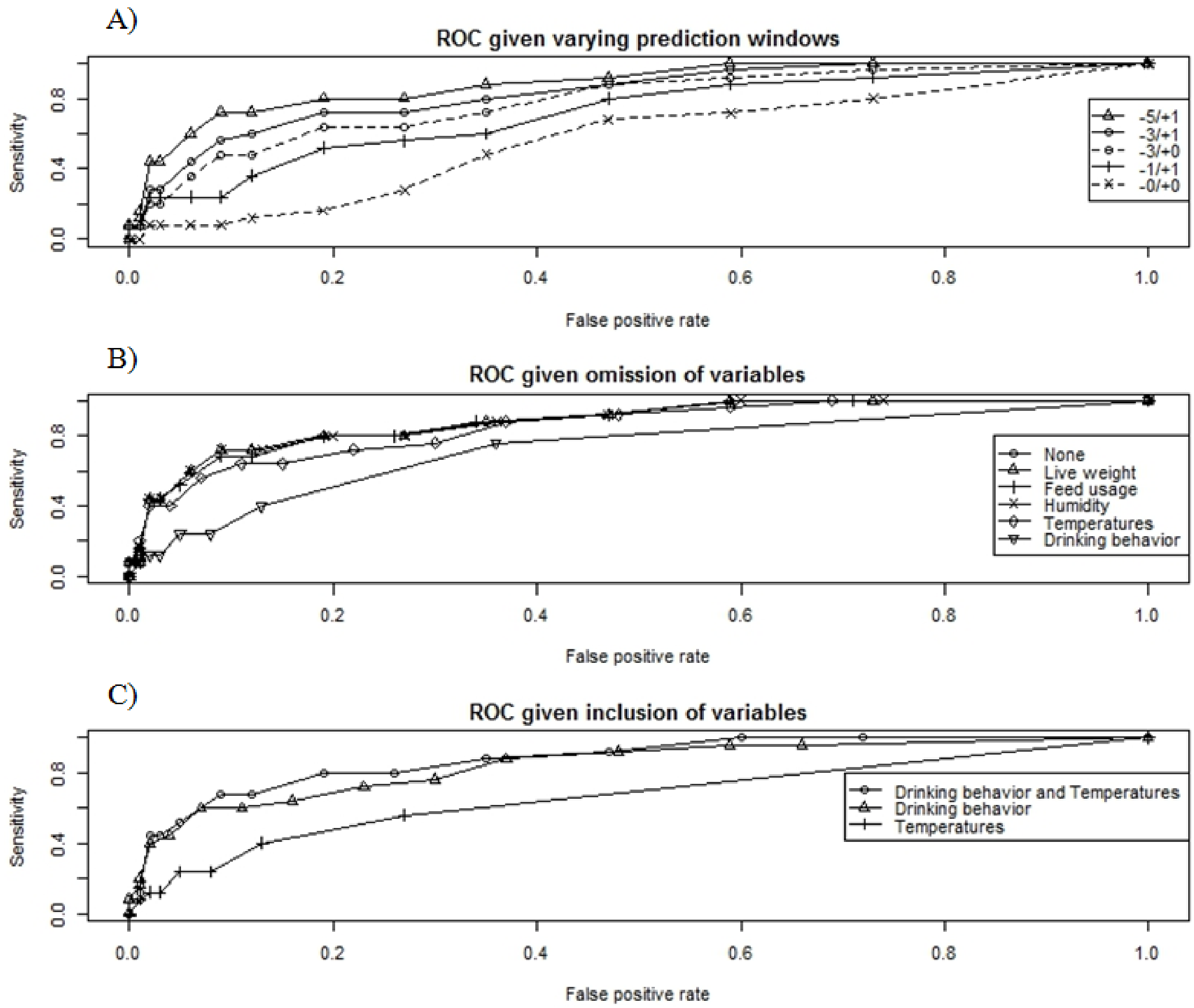



**Figure 10: A) ROC curves given five selected time windows when all seven variables are included. B) ROC curves given the omission of selected variables. Drinking behavior refers to both water flow and drinking bouts frequency. Temperatures refer to both back wall and corridor temperatures. C) ROC curves achieved when only selected variables (temperate and/or drinking behavior) are included.**

**Table 10: The predictive performance achieved when omitting the various variables**

| Omitted variables | AUC |
|---|---|
| None | 0.88 |
| Mean live weight | 0.88 |
| Mean feed usage | 0.88 |
| Humidity | 0.88 |
| Temperature, corridor | 0.88 |
| Temperature, drinking nipple | 0.85 |
| Temperatures, corridor & drinking nipple | 0.85 |
| Water flow | 0.85 |
| Drinking bouts frequency | 0.87 |
| Water flow & drinking bouts frequency | 0.73 |

Based on the fact that temperature and drinking behavior are the only two variables seen to affect the predictive performance, it was decided to test what performances could be achieved when including only both of these as well as each of these alone. It was further decided to test these values with both a -5/+1 window and a -3/+1 window. The -5/+1 window shows what performance can maximally be expected given the DLM/Cholesky method using the model described in this paper. The -3/+1 window allows the best possible comparison between the performance of the DLM/Cholesky method with the temperature-only summary/logistic regression method presented in a previous study (Jensen & Kristensen 2016). All results are seen in Table 11, while the ROC curves for the -5/+1 windows are seen in Figure 10C. Notice first that for the -3/+1 window, the AUC produced when including both temperature variables and drinking behavior is 0.82, compared to the 0.83 seen in Table 9. This means that when using this prediction window, the combined omission of live weight, feed usage, and humidity amount to a reduction of 1 percentage point of the AUC. When using the -5/+1 window, however, this combined effect is not seen.

Secondly, only including temperature reduces the performances by between 21 and 18 percentage points, *i.e.* to 0.67 and 0.65, given the -5/+1 and the -3/+1 prediction windows, respectively. It should be noted that the AUC of 0.67 is considerably lower than the AUC of 0.73, as seen in Table 10 when both water flow and drinking bouts frequency were excluded. This shows that live weight, feed amount, and/or humidity actually do add some information when drinking behavior data is not available, but that this information value disappears completely when drinking behavior is available. Further testing showed that all of this information was primarily contained in feed usage, secondarily in humidity, and that live weight added no information (data not shown).
Only including drinking behavior reduces the AUC by a mere 2 and 4 percentage points, compared to including all variables.

**Table 11: The predictive performances given the inclusion of only drinking behavior (water flow and drinking bouts frequency), temperature measurements (by the back wall and by the corridor), or both drinking behavior and temperature measurements.**

| Included variables | Prediction window | AUC |
|---|---|---|
| Drinking behavior; Temperatures, corridor and back wall | -5/+1 | 0.88 |
| | -3/+1 | 0.82 |
| Temperatures, corridor and back wall | -5/+1 | 0.67 |
| | -3/+1 | 0.65 |
| Drinking behavior | -5/+1 | 0.84 |
| | -3/+1 | 0.81 |

Together, these results paint a clear picture that the drinking behavior contain the most information in relation to detection and early warnings of diarrhea and pen fouling. The observation that drinking behavior of the pigs is significantly altered when undesired events, such as diarrhea and pen fouling, are under way, are in accordance with the findings of Madsen and Kristensen (2005). Furthermore, the fact that drinking bouts frequency, which served as our proxy for pig activity, can be seen to contain information independently of water flow, suggests that better and more direct automatic observations of the activity of the pigs would be worth investigating. This could be done, for example, via cameras above the pen, combined with some activity analyzing software.

It is worth pointing out that in our previous study (Jensen & Kristensen 2016), an AUC of 0.80 was achieved for indiscriminate warnings of diarrhea and pen fouling when only using temperature summaries and a -3/+1 prediction window. In this study, using the same window and including all available variables, we achieved an AUC of 0.83, while the performance fell to an AUC of 0.65 when only temperature data were included. This shows rather convincingly that while the DLM/Cholesky method used here may be well suited for harnessing the information in the drinking behavior of the pigs, a per-day summary approach is the better choice for the temperature data. The challenge of combining these rather different types of monitoring methods will thus have to be addressed in later research.

It is also worth noting that the variables with the least effect on performance were the variables with the lowest frequency of observations, namely live weight (observed weekly), feed amount (observed daily) and humidity (observed daily, sometimes hourly). It is conceivable that a more continuous monitoring of the animals live weight would increase the relative value of the weight information, as deviations from expected growing patterns could be detected sooner. This could for example be achieved via video monitoring of the pen combined with software capable of estimated the average weight of the pigs based on their observed size.

It might further be that the relative value of the feed amount and humidity observations would be greater, if the data were modeled with time steps of one day rather than one hour. Alternatively, like the temperature data, these variables may need to be monitored in a completely different way from what has been done here in order to provide any useful information. Further studies would be needed to evaluate these strategies.

Furthermore, it is worth noticing some interesting differences between the data observed on days with diarrhea compared to the data observed on days with pen fouling, as seen in Table 1. First of all, diarrhea is associated with higher water flow and water flow frequencies during the night hours, *i.e.* between 9 PM and 9 AM. This same association is not seen with pen fouling. Diarrhea is also associated with lower corridor temperatures, compared to days with no events, whereas pen fouling is associated with higher corridor temperatures. Pen fouling is further associated with greater diurnal changes in temperature at the back wall, and higher overall temperatures at the corridor. Remember that the DLM/Cholesky method presented in this paper is not able to distinguish between the two types of events, but merely raise an alarm whenever the system is changing compared to the healthy situation. Thus, if the goal is early warning and prevention, a response targeted at a specific event would be preferable, as different responses to the two problems would be required; in case of oncoming diarrhea, the proper response would be to verify that the pigs have an intestinal infection and then treat with the appropriate antibiotics, while in the case of pen fouling, the proper response would be to regulate the temperature or to lower the stocking density in the pen. Given the differences in the nature of the two events included in this study, as described above, making a system to raise specific alarms about a specific events being under way could conceivably be achieved. This could be done by, for example, not unifying the forecast errors with the Cholesky method used here, but by instead parsing the forecast errors to *e.g.* a Bayesian classifier or an artificial neural network.

Lastly, it should be remembered that only registrations of diarrhea and pen fouling were actually available for this study, yet it is known that other undesired events such as pneumonia and influenza did occur. For these reasons, the specificities reported in this paper are likely to be underestimated.

# 4 CONCLUSIONS

We demonstrate the use of a multivariate dynamic linear model as a method for combining multiple and diverse precision data sensors, where data with different origin, numerical magnitude, and observational frequency can be co-modeled, and the forecast errors of this model translated into alarms concerning undesired events. This framework of DLM-based alarms further provide an easy method of estimating the relative information value of the various data streams by systematically omitting and including single data streams in the model.

Moreover, we show that early but indiscriminate warnings can be raised for diarrhea and pen fouling using the unified forecast errors of the multivariate dynamic linear model using time steps of one hour. The best performance was achieved with a prediction window of 7 days (-5/+1 days, relative to the assigned event observation), with an area under the receiver operating characteristics curve of 0.88 and a specificity of 0.81 when the sensitivity is held at 0.80. Using longer prediction windows than this did not improve the predictive performance. Drinking behavior, *i.e.* water flow and drinking bouts frequency, were found to be by far the most significant predictors of these undesired events, followed by temperature measurements.

## 5 ACKNOWLEGDEMENTS

This research was carried out with support from The Danish Council for Strategic Research (The PigIT Project, Grant number 11-116191). We further wish to thank the anonymous farmer, the Danish Pig Research Center and the technical staff at Aarhus University for installing and supervising the sensors and for taking care of the daily observations in the herd. A special thanks to Mads Ravn Jensen, who designed, installed, and calibrated the sensors which were used for monitoring the drinking behavior of the pigs in this study.

## 6 REFERENCES

Hogeveen, H. et al., 2010. Sensors and clinical mastitis--the quest for the perfect alert. *Sensors (Basel, Switzerland)*, 10(9), pp.7991–8009.

Jensen, D.B. et al., 2015. A multi-dimensional dynamic linear model for monitoring slaughter pig production. In *7th European Conference on Precision Livestock Farming*. pp. 503–512.

Jensen, D.B. & Kristensen, A.R., 2016. Temperature as a predictor of fouling and diarrhea in slaughter pigs. *Livestock Science*, 183, pp.1–3.

Jensen, D.B., Toft, N. & Cornou, C., 2014. The effect of wind shielding and pen position on the average daily weight gain and feed conversion rate of grower/finisher pigs. *Livestock Science*, (167), pp.1–9.

Lopez, J. et al., 1991. Effects of temperature on the performance of finishing swine : II . Effects of a cold , diurnal temperature on average daily gain , feed intake , and feed efficiency. *Journal of animal science*, 69, pp.1850–1855.

Madsen, T.N., Andersen, S. & Kristensen, A.R., 2005. Modelling the drinking patterns of young pigs using a state space model. *Computers and Electronics in Agriculture*, 48(1), pp.39–62.

Madsen, T.N. & Kristensen, A.R., 2005. A model for monitoring the condition of young pigs by their drinking behaviour. *Computers and Electronics in Agriculture*, 48(2), pp.138–154.

Montgomery, D., 2005. *Introduction to Statistical Quality Control* 5th ed., Hoboken, NJ, USA: Wiley.

Ostersen, T., Cornou, C. & Kristensen, A.R., 2010. Detecting oestrus by monitoring sows' visits to a boar. *Computers and Electronics in Agriculture*, 74(1), pp.51–58.

Stege, H. et al., 2011. Association between lean meat percentage and average daily weight gain in Danish slaughter pigs. *Preventive Veterinary Medicine*, 101(1-2), pp.121–123.

The R Core Team, 2013. *R : A Language and Environment for Statistical Computing* 2.15.3 ed., R Foundation for Statistical Computing.

West, M. & Harrison, J., 1997. *Bayesian Forecasting and Dynamic Models* 2nd ed., New York, USA: Springer.

Young, B.A., 1981. Cold stress as it affects animal production. *Journal of Animal Science*, (1), pp.154–163.

Zweig, M.H. & Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4), pp.561–577.

Aarnink, A.J.A. et al., 2006. Temperature and body weight affect fouling of pig pens. *Journal of Animal Science*, 84, pp.2224–2231.

# Chapter 9: Paper 5

Manuscript, submitted for review

**Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis**

**Dan B. Jensen*[1], Henk Hogeveen†, and Albert De Vries‡**
* Department of Large Animal Sciences, University of Copenhagen, DK-1870 Frederiksberg C, Denmark
† Chair group Business Economics, Wageningen University, 6706 KN Wageningen, the Netherlands
‡ Department of Animal Sciences, University of Florida, Gainesville 32611

[1] Corresponding author: daj@sund.ku.dk

## ABSTRACT

Rapid and an automatic detection of dairy cow mastitis is important so corrective action can be taken as soon as possible while labor needs for monitoring are reduced. Automatically collected sensor data used to monitor the performance and the health state of the cow could be useful for automatic detection of mastitis. The state of the art in combining sensor data to predict clinical mastitis still does not yield sufficient performance to be applied in practice. Our objective was to present the combination of a multivariate dynamic linear model (DLM) with a naïve Bayesian classifier (NBC) as a novel method for combining sensor- and non-sensor data for the purpose of detecting clinical cases of mastitis. We also evaluated reductions in the number of sensors for detecting mastitis. With the DLM we co-modeled 7 sources of sensor data (milk yield, fat, protein, lactose, conductivity, blood, body weight) collected at each milking for individual cows to produce one-step-ahead forecasts for each sensor. The observations were subsequently categorized according to the errors of the forecasted values and the estimated forecast variance. The categorized sensor data was combined with other data pertaining to the cow (week in milk, parity, mastitis history, SCC category, and season) using Bayes' theorem which produced a combined probability of the cow having clinical mastitis. If this probability was above a set threshold, the cow was classified as mastitis positive. To illustrate the performance of our method, we used sensor data from 1,003,207 milkings from the University of Florida Dairy Unit collected from 2008 to 2014. Of these, 2,907 milkings were associated with recorded cases of clinical mastitis. With this DLM/NBC method, we reached an area under the receiver operating characteristic curve of 0.89, with a specificity of 0.81 when the sensitivity was set at 0.80. Specificities with omissions of sensor data ranged from 0.58 to 0.81. These results were comparable to other studies, but differences in data quality, definitions of clinical mastitis, and time windows make comparisons across studies difficult. We found the DLM/NBC method to be a flexible method to combine multiple sensor and non-sensor data sources to predict clinical mastitis and accommodate missing observations. The DLM method produces forecasts that are approximately normally distributed, which makes forecasts and forecast errors easy to interpret and new sensors can easily be added

**Key Words:** mastitis, Bayesian classifier, dynamic linear model

## 1 INTRODUCTION

Mastitis is associated with a wide range of characteristics that can be measured in milk. In a classic review, Kitchen (1981) described the effect of mastitis on the composition of milk and discussed potential diagnostics based upon these effects. In addition to SCC, electrical conductivity, milk constituents (especially lactose) and enzymes (such as N-acetyl-β-D-glucosaminidase and lactate dehydrogenase) have been identified to be affected by clinical mastitis.

Since the 1990s, work has been carried out on automated detection of mastitis using changes in one or more milk characteristics (e.g. Nielen et al., 1996). Automated mastitis detection systems started to be widely used on commercial dairy farms with the introduction of automatic milking systems approximately 20 yr ago. A mastitis detection system consists of at least 2 elements: the sensor (hardware), and the algorithms to translate sensor data into alerts (software). A decision support system and a decision making system are possibly also part of a mastitis detection system (Rutten et al., 2013). The main sensor (hardware) that is being used to detect mastitis measures electrical conductivity (e.g., Nielen et al., 1995, Norberg et al., 2004; Cavaro et al., 2006). Sensor systems based on other milk characteristics are also proposed and on the market, such as milk color (Song et al., 2010), lactate dehydrogenase (Chagunda et al., 2006; Friggens et al., 2007) and SCC (Whyte et al., 2005; Mollenhorst et al., 2010).

Most publications on automated mastitis detection systems are aimed at the algorithm to transform sensor data into alerts. Quite some different data modeling techniques have been proposed, including thresholds (e.g., Mollenhorst et al., 2010), moving averages (Maatje et al., 1992), neural networks (e.g., Nielen et al., 1992; Cavero et al., 2008), fuzzy logic (e.g., de Mol and Woldt, 2001; Kamphuis et al., 2008), time series analysis (de Mol, 2001; Cavero et al., 2007), discriminant function analysis (Norberg et al., 2004); Kamphuis et al., 2010), and wavelet filtering (Miekley et al., 2013). In most of these studies, electrical conductivity was combined with other measurements (mostly with milk yield) to improve the performance of the detection system.

So far, the performances of the published mastitis detection systems do not satisfy the high accuracy needed for practical clinical mastitis detection systems (Hogeveen et al., 2010). Combining data from more sources has been suggested as a possible method to improve the performance of mastitis detection systems. It is still unclear how to best combine data from different sensors and other data, including accounting for missing observations. Bayesian analysis has been used as an approach to prioritize sensor data-based alarms by including cow specific information (Steeneveld et al., 2010).

Most mastitis detection systems compare observed sensor values to forecasted values and monitor forecast errors. Forecasts are typically based on moving averages (e.g. Maatje et al., 1992) but if the quality of the forecast is improved, then the performance of a mastitis detection system may be improved as well.

As a method for combining the many possible lines of sensor- and non-sensor based data for a unified prediction of mastitis, we propose using a multivariate dynamic linear model (DLM) in combination with a naïve Bayesian classifier (NBC). The multivariate DLM provides the forecast values while the NBC combines all available observations, including forecast errors, with a prior probability to achieve a single posterior probability of mastitis.

A property of the multivariate DLM, as described by West and Harrison (1997), is that it is adaptive, and thus the expected values are automatically adjusted to the longer term trend of the data. Another property of the multivariate form of the DLM is that the co-dependencies between several variables of interest can be taken into account, when one-step-ahead forecasts for these variables are calculated, which is attractive for the NBC.

Similar adaptive forecasting has been applied by Huybrechts et al. (2014), who used a synergistic control process to adjust lactation curves in an effort to use milk yield as a predictor of clinical mastitis (sensitivity: 0.63). Huybrechts et al. (2014) relied heavily on a specific mathematical model for long term forecasting, whereas the adaptive and short-term nature of the forecasts produced by a DLM allows for a freer description of multiple (non)-linear trends that may predict the short-term observations better. Furthermore, the DLM easily handles missing data, as one-step-ahead forecasts are always produced given the available data.

Few applications of DLM for monitoring animal production systems exist. Univariate implementations of the DLM have been developed for applications including detection of estrus in sows (Ostersen et al., 2010) and describing the drinking behavior of young pigs (Madsen and Kristensen, 2005). To our knowledge, no previous descriptions of applications of a multivariate DLM exist with the purpose of detecting diseases in production animals, such as mastitis in dairy cows.

An NBC classifies a new set of observations by estimating the probability that the observation belongs to each class (mastitis or healthy). The NBC is a relatively simple classification method, but it has been shown to be useful in a wide range of fields, such as prediction of bacterial thermophilicity (Jensen et al., 2012), diagnosis of classical swine fewer (Geenen et al., 2011) and detection of clinical mastitis (Steeneveld et al., 2009). The NBC has advantages over comparable classification methods, such as artificial neural networks or logistic regression functions, because missing observations can be easily handled in an NBC by including only those observations which are available. Similarly, adding data from a new sensor is relatively trivial with the NBC, so long as likelihoods are available for the outputs of that sensor, associated with the outcome variable that needs to be classified. Such likelihoods may be estimated from scientific literature or practical knowledge of how mastitis influences milk characteristics and cow physiology, or they may be derived directly from observations made on-site using the sensor. Lastly, the likelihoods make it easy to see the relative contributions of the various variables of interest, as opposed to the black box-nature of for example neural networks.

A combination of a DLM and NBC is therefore a potentially attractive practical method to detect clinical mastitis using data from multiple sources. We have 2 objectives with this study: 1) to describe and illustrate the combination of a multivariate DLM and a NBC for detecting clinical mastitis; 2) to measure the performance of the DLM/NBC method and estimate the relative importance of various combinations of sensor- and non-sensor data on that performance.

## 2 MATERIALS AND METHODS

### 2.1 Data Sources

In this study, we refer to 2 types of data, namely continuous data and categorical data. Continuous data are those which were obtained with sensors and used by the DLM in their raw, numeric form. Categorical data are those which are considered to fall within 2 or more separate categories, regardless of whether the information was collected using sensors or not.

All data were obtained from the University of Florida Dairy Unit located in Hague, Florida, between September 2008 and March 2014. The herd consisted of approximately 500 Holstein cows which were housed in freestall barns with sand bedding and fans and sprinklers for heat abatement. Cows were fed a total mixed ration and milked twice per day with 12-h intervals in a double-12 milking parlor.

In and around the milking parlor, milk yield, milk conductivity, fat %, protein %, lactose %, blood %, SCC category and body weight were automatically recorded per cow per milking using sensors. Blood % is the volume of blood in 1 ml of milk expressed as a percentage. Milk yield and conductivity were measured with milk meters. Milk components and SCC categories were measured by real-time milk analyzers (AfiLab). Body weights were measured by automated scales when cows exited the milking parlor. All sensors were obtained from Afimilk, Kibbutz Afikim, Israel. The accuracies of the sensors for fat, protein and lactose have been evaluated by Kaniyamattam and De Vries (2014). The 7 continuous variables from these sensor data were used in the DLM.

The 5 categorical variables were SCC category (0-200, 200-400, 400-800, 800+ x 1,000 cells/ml), parity (first, later), previous mastitis treatment (yes, no, excluding treatment at the time of observation), season (warm: May to August; cold: September to April), and the week in milk (WIM). These 5 categorical variables were not used in the DLM but added in the NBC.

All records of clinical mastitis cases were obtained from the herd management information system. Clinical mastitis cases were determined in the milking parlor by trained farm staff following the farm's standard operating protocols (Donovan et al., 2011) which included forestripping and visual observation of milk. In addition, cows on alert lists for larger than expected milk yield deviations were evaluated for clinical mastitis by farm staff.

If mastitis was diagnosed during the evening milking, the cow was immediately moved after her milking to the hospital herd (but not treated) and then evaluated and treated after her next milking in the morning. If mastitis was diagnosed during the morning milking, the cow was immediately moved to the hospital herd after her milking. The hospital herd was the last group milked in each milking shift. Treatment of mastitis occurred in the morning when the hospital herd was milked. Diagnosis and treatment were conducted by trained farm staff or veterinarians of the University of Florida. Confirmed clinical mastitis cases were treated according the standard operating protocol and entered in the management information system.

All mastitis diagnoses were registered on a daily level, and the milking of first diagnosis was not available. For example, if mastitis was registered on a Friday, the actual mastitis diagnosis was made either during the Friday morning milking or the Thursday evening milking. Such daily mastitis recording is routine in the US because the aim is to track withholding times for meat and milk after treatment.

### 2.2 Data Editing

All data editing, modeling and calculations were done using the statistical language and environment R (The R Core Team, Austria 2013).

The sensor data were screened for outliers using simple summary statistics. We judged all observations from the 7 continuous sensor variables to be within acceptable ranges.

Data collected for a given cow within at least 14 days of a new mastitis observation were not considered when calculating the updated probability of mastitis, and such data are not included in the following data descriptions.

All continuous sensor variables and SCC categories contained some missing values due to automatic data reading or entry problems in the parlor. The most extreme cases where body weight and SCC category, with a total of 157,250 (15.67%) and 61,429 (6.12%) of these observations missing, respectably. The remaining sensor variables had between 15,595 (1.55%) and 31,071 (3.10%) missing observations. Often the missing data occurred in the same milkings. The periods of missing data were on average between 1.32 and 1.84 observations long for the individual variables, but in rare cases the periods for individual missing variables were up to 257 observations. A total of 9 out of 2,051 lactations had periods of missing data for more than 100 consecutive observations for at least one sensor variable. None of the other categorical variables (*i.e.* season, parity, previous mastitis treatments, and WIM) had missing observations.

New clinical mastitis cases were associated with both the morning and evening before milking because we could not distinguish if the case was first diagnosed in the evening or in the morning. Any case recorded within 14 d (28 milkings) of a new case was considered to be a flare up of the same case for that cow and was ignored.

The descriptive statistics of the edited data are shown in Table 1. Data from 1,003,207 milkings were available, including 2,907 milkings (morning and evening) from days where clinical mastitis was recorded. Figure 1 shows the number and prevalence of mastitis observations in the first 43 WIM.

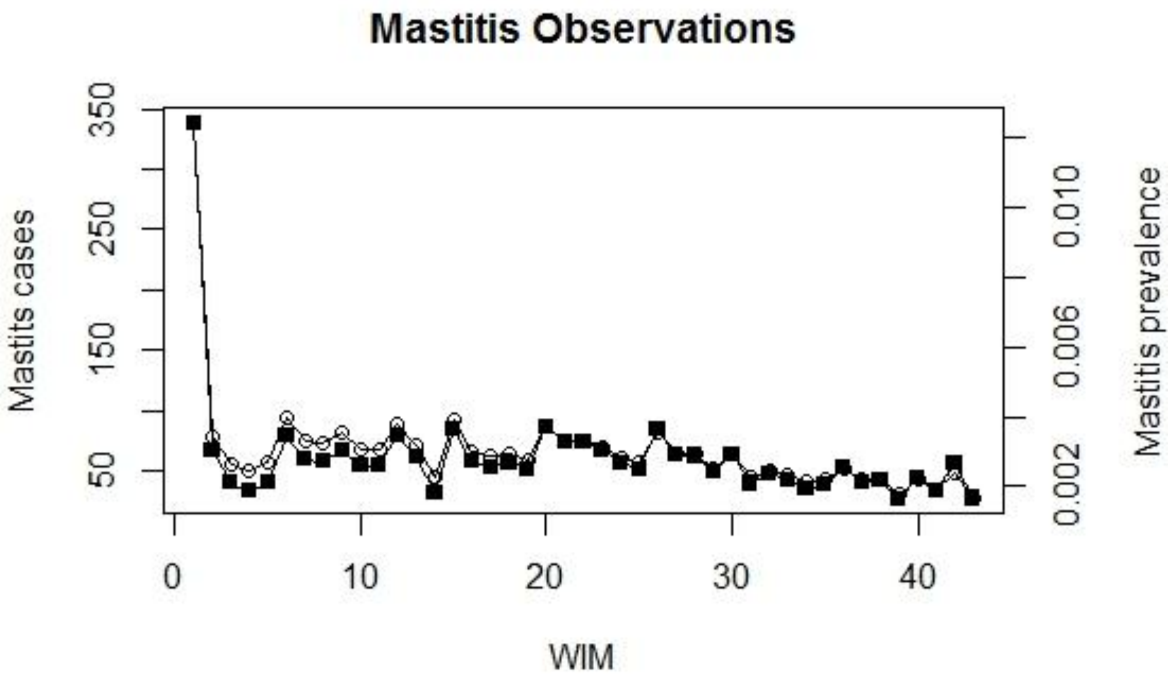



**Figure 1: The number of mastitis cases (circles; left vertical axis) and the prevalence of mastitis cases (solid squares; right vertical axis) observed for the entire study population, *i.e.* all cows from the learning set and test set, by week in milk (WIM).**

**Table 1. Descriptive statistics of the study data collected at the University of Florida Dairy Unit. Data collected for a given cow within at least 28 milkings of a mastitis observation are not included.**

| Categorical variables | N[1] | Outcome | Continuous variables | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | mastitis cases | milk yield (kg) | | milk conductivity | | fat (%) | | protein (%) | | lactose (%) | | blood (%) | | body weight (kg) | |
| | | | mean | SD | mean | SD | mean | SD | mean | SD | mean | SD | mean | SD | mean | SD |
| Total | 1,003,207 | 2,907 | 16.01 | 5.12 | 10.93 | 1.35 | 3.67 | 0.53 | 3.11 | 0.28 | 4.76 | 0.28 | 0.22 | 0.09 | 625.05 | 85.23 |
| Time of day | | | | | | | | | | | | | | | | |
| Morning | 502,335 | 1,464 | 16.61 | 5.23 | 11.03 | 1.35 | 3.64 | 0.53 | 3.10 | 0.28 | 4.77 | 0.29 | 0.22 | 0.10 | 616.58 | 83.90 |
| Evening | 500,872 | 1,443 | 15.42 | 4.93 | 10.83 | 1.35 | 3.70 | 0.53 | 3.11 | 0.27 | 4.76 | 0.28 | 0.21 | 0.09 | 633.28 | 85.72 |
| Current mastitis | | | | | | | | | | | | | | | | |
| Yes | 2,907 | 2,907 | 13.09 | 7.27 | 12.43 | 2.05 | 4.02 | 0.88 | 3.24 | 0.50 | 4.39 | 0.57 | 0.29 | 0.13 | 638.17 | 85.13 |
| No | 1,000 300 | 0 | 16.02 | 5.11 | 10.92 | 1.35 | 3.67 | 0.53 | 3.11 | 0.28 | 4.77 | 0.28 | 0.22 | 0.09 | 624.98 | 85.23 |
| Season | | | | | | | | | | | | | | | | |
| Cold | 665,457 | 1,625 | 16.06 | 5.19 | 10.94 | 1.38 | 3.70 | 0.54 | 3.11 | 0.28 | 4.78 | 0.30 | 0.21 | 0.09 | 622.22 | 84.57 |
| Warm | 337,750 | 1,282 | 15.92 | 4.97 | 10.91 | 1.30 | 3.63 | 0.51 | 3.10 | 0.28 | 4.74 | 0.25 | 0.22 | 0.10 | 630.49 | 86.24 |
| SCC category[2] | | | | | | | | | | | | | | | | |
| 0-200 | 768,059 | 1,152 | 16.53 | 4.83 | 10.86 | 1.26 | 3.62 | 0.47 | 3.07 | 0.23 | 4.85 | 0.16 | 0.21 | 0.09 | 621.67 | 83.65 |
| 200-400 | 82,240 | 392 | 14.64 | 5.16 | 11.22 | 1.59 | 3.85 | 0.55 | 3.19 | 0.26 | 4.61 | 0.18 | 0.23 | 0.10 | 646.95 | 89.27 |
| 400-800 | 34,342 | 240 | 14.08 | 5.38 | 11.23 | 1.65 | 3.88 | 0.62 | 3.24 | 0.30 | 4.48 | 0.21 | 0.24 | 0.10 | 645.07 | 91.36 |
| 800+ | 57,137 | 563 | 13.48 | 5.66 | 11.24 | 1.65 | 3.88 | 0.73 | 3.32 | 0.40 | 4.21 | 0.31 | 0.27 | 0.11 | 638.85 | 91.58 |
| Parity | | | | | | | | | | | | | | | | |
| First | 421,606 | 611 | 14.19 | 3.88 | 10.58 | 1.12 | 3.64 | 0.50 | 3.10 | 0.26 | 4.79 | 0.28 | 0.21 | 0.09 | 568.71 | 64.59 |
| Later | 580,396 | 2 314 | 17.32 | 5.50 | 11.18 | 1.45 | 3.70 | 0.55 | 3.11 | 0.29 | 4.75 | 0.29 | 0.22 | 0.10 | 664.89 | 74.96 |
| Previous mastitis[3] | | | | | | | | | | | | | | | | |
| Yes | 356,866 | 1 689 | 16.18 | 5.26 | 11.23 | 1.49 | 3.70 | 0.53 | 3.11 | 0.27 | 4.74 | 0.27 | 0.23 | 0.10 | 662.19 | 81.66 |
| No | 646,341 | 1 218 | 15.92 | 5.04 | 10.77 | 1.24 | 3.66 | 0.53 | 3.10 | 0.28 | 4.78 | 0.29 | 0.21 | 0.09 | 604.71 | 80.14 |

[1] Number of milkings
[2] SCC is listed as thousands of cells/ml
[3] Binary variable, becomes Yes after a cow has completed its first mastitis treatment

The available data were split evenly into a learning set (500,442 milkings, 1,455 of which were on days with mastitis diagnosis) for estimating parameters for the DLM and the likelihoods for the NBC, and a test set (502,765 milkings, 1,452 of which were on days with mastitis diagnosis) for validation of the method. This was done by randomly assigning cows to the learning or test set and subsequently writing all observations from all lactations of the selected cows to their assigned set. This ensured that the observations in the 2 sets were as independent of each other as possible.

### 2.3 Application of Multivariate Dynamic Linear Models

Multivariate dynamic linear models with one step Markov evolution (West and Harrison, 1997) were used to forecast the continuous sensor values for each individual cow at each milking. Because the mean level at any given DIM was clearly different between morning and evening milkings (Palmer et al., 1994) for several of the sensors, as indicated in Table 1, the morning and evening data were modeled separately. Thus each DLM model would continuously forecast the 7 continuous sensor observations for the next morning or the next evening milking.

We expand on these methods, as they are relatively unknown in the dairy science community. In general, a DLM consists of an observation equation and a system equation (equations 1 and 2, respectively) as follows:

$$\boldsymbol{Y_t} = \boldsymbol{F_t'}\boldsymbol{\theta_t} + \boldsymbol{v_t}, \qquad \boldsymbol{v_t} \sim N\left(\underline{0}, \boldsymbol{V_t}\right) \tag{1}$$

$$\boldsymbol{\theta_t} = \boldsymbol{G_t}\boldsymbol{\theta_{t-1}} + \boldsymbol{w_t}, \qquad \boldsymbol{w_t} \sim N\left(\underline{0}, \boldsymbol{W_t}\right) \tag{2}$$

Equation 1 describes how the values of an observation vector ($\boldsymbol{Y_t}$) depend on an unobservable parameter vector ($\boldsymbol{\theta_t}$) to time $t$. To describe both level and trend for each sensor, the parameter vector ($\boldsymbol{\theta_t}$) contains the underlying values for each of the continuous sensor variables, as well as the trend of the variable, *i.e.* the rates at which those same values change at time $t$. The initial mean levels for each variable were set as the mean value of the first observations of the respective variables across all lactations in the learning set. The initial trend was estimated as the average change in mean level from DIM = 1 to DIM = 2. These initial means and trends were estimate separately for morning and evening observations.

The system matrix ($\boldsymbol{G_t}$ of equation 2) serves to update the expected values of the observable variables from time $t-1$ to time $t$ by adding the trend to the current level at that time. The transposed design matrix $\left(\boldsymbol{F_t'}\right)$ serves to extract the expected values of the observable variables from the parameter vector, thus yielding a vector which includes only these estimates.

In our case, the system matrix as well as the 2 variance matrices are constant, so that $G_t = G$, $V_t = V$, and $W_t = W$. $F_t$ varies over time because it depends on which of the 7 sensor variables have missing observations at a given time, as explained in the multivariate case below.

### 2.3.1 Univariate Example

Assume that we wish to model the morning milk yield alone with a univariate DLM. We estimate that the initial morning milk yield on day 1 is 0.47 kg and that the initial trend in morning milk yield is +0.21 kg/d. Thus we can describe the change in morning milk yield from DIM = 1 to DIM = 2 according to equation (2) as follows:

$$\boldsymbol{\theta_2} = \boldsymbol{G} \cdot \boldsymbol{\theta_1} + \boldsymbol{w}$$

$$\boldsymbol{\theta_2} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.47 \\ 0.21 \end{bmatrix} + \boldsymbol{w} \tag{2.e}$$

$$\boldsymbol{\theta_2} = \begin{bmatrix} 0.68 \\ 0.21 \end{bmatrix} + \boldsymbol{w}$$

We can further describe the estimated milk yield observation at DIM = 2 according to Equation (1) as follows:

$$\boldsymbol{Y_2} = \boldsymbol{F_t'} \cdot \boldsymbol{\theta_2} + \boldsymbol{v}$$

$$\boldsymbol{Y_2} = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0.68 \\ 0.21 \end{bmatrix} + \boldsymbol{v} \tag{1.e}$$

$$\boldsymbol{Y_2} = 0.68 + \boldsymbol{v}$$

In other words, the prior estimates of the expected observations to a given time are calculated using the common rules for matrix multiplication.

### 2.3.2 The Multivariate Case

Constructing the multivariate DLM is accomplished by combining the univariate models needed to describe the individual variables, as shown in the example above, while taking into account the co-variances between those variables. Here the transposed design matrix $(\boldsymbol{F_t'})$ and the system matrix $(\boldsymbol{G})$ have repeated structures corresponding to the number of variables being modeled, as illustrated by equations 3 and 4, respectively.

$$\boldsymbol{F_t'} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \tag{3}$$

$$\boldsymbol{G_t} = \begin{bmatrix} 1 & 1 & \cdots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & & 0 & 1 \end{bmatrix} \tag{4}$$

When all 7 sensor variables are included, $\boldsymbol{F_t'}$ has 7 rows and 14 columns, while $\boldsymbol{G}$ has 14 rows and 14 columns. As seen in equation 4, $\boldsymbol{G}$ is a block-diagonal matrix, where each 2x2 block serves to update the expected value of a specific sensor variable, by adding the appropriate trends to the current values, as in the univariate example above. Also in accordance with the univariate example, the structure of $\boldsymbol{F_t'}$ serves to separate those same observable values from the unobservable trends, yielding a vector of expected observations to each time, $t$.

For $\boldsymbol{F_t'}$, only those rows corresponding to the variables which are actually observed at a given time are included at that time. This ensures that missing observations are ignored when the parameter vector is updated through Kalman filtering (West and Harrison, 1997). Kalman filtering is a method for reducing the noise in the observed data by considering the difference between the observed and forecasted values (i.e. the forecast error), while taking into consideration the variances associated with forecasting and observations.

The co-dependencies between the various observable variables (1 to $n$, with $n$ = 7 in this study), as well as their unobservable trend values ($d1$ to $dn$) are accounted for by the observational co-variance matrix ($\boldsymbol{V}$) and the system evolution co-variance matrix ($\boldsymbol{W}$), as illustrated by equations 5 and 6, respectively.

$$\boldsymbol{V} = \begin{bmatrix} V_{1,1} & \cdots & V_{1,n} \\ \vdots & \ddots & \vdots \\ V_{n,1} & \cdots & V_{n,n} \end{bmatrix} \tag{5}$$

$$\boldsymbol{W} = \begin{bmatrix} W_{1,1} & W_{1,d1} & \cdots & W_{1,n} & W_{1,dn} \\ W_{d1,1} & W_{d1,d1} & \cdots & W_{d1,n} & W_{d1,dn} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ W_{n,1} & W_{n,d1} & \cdots & W_{n,n} & W_{n,dn} \\ W_{dn,1} & W_{dn,d1} & \cdots & W_{dn,n} & W_{dn,dn} \end{bmatrix} \tag{6}$$

For example, if milk yield is considered variable number 1 and electrical conductivity is variable number 2, then $V_{1,1}$ is the observational variance of milk yield, $V_{2,2}$ is the observational variance of electrical conductivity, and $V_{1,2}$ is the observational co-variance between milk yield and electrical conductivity. Similarly, $W_{1,1}$ is the systematic variance of the evolution of milk yield, $W_{1,d1}$is the systematic co-variance between milk yield evolution and the evolution of the trend (rate of change) of milk yield and $W_{1,2}$ is the systematic co-variance between the evolution of milk yield and electrical conductivity. Thus the co-variances found for the off-diagonal positions of the $\boldsymbol{V}$and $\boldsymbol{W}$ matrices are what provides the extra information about variable interaction, compared to the information available in the univariate models.

We estimated the values in the $\boldsymbol{V}$ and $\boldsymbol{W}$ matrices using the expectation maximization (EM) algorithm (West and Harrison, 1997), applied to the learning set. The variances of the different variables were adjusted to similar scales by dividing all milk yield observations by 10 and all body weight observations by 100 before modeling. To ensure that the DLM was optimized for modeling the healthy, non-mastitic, state of the cows, only lactations with at least 43 full WIM, with at most 10 % missing data overall, and missing data periods of at most 2 consecutive observations, were used (N = 41 lactations) from the learning set. Convergence of the EM algorithm was determined by plotting variance values, *i.e.* the diagonal values of V and W, against the number of iterations, and inspecting these plots visually. Convergence was reached when the EM algorithm was run for 300 iterations for the morning model and 230 iterations for the evening model.

To demonstrate how the multivariate DLM's ability to accurately forecast the continuous variables was affected by the presence of mastitis, the normalized forecast errors from the test set were selected for all morning milkings on days with and without mastitis, separately. For each continuous variable, the mean for non-mastitis forecast errors were calculated for each DIM. These healthy-associated mean errors were then plotted along with the individual forecast errors made for the same sensor variable, on the days where the cow was positive for clinical mastitis.

### 2.4 Learning Likelihoods for Bayesian Classification

The forecast errors of a multivariate DLM are normally distributed, and any interdependencies between data from the 7 sensors are accounted for by the co-variance matrices. This makes them suitable input parameters for an NBC. The DLM was run from DIM = 1 to DIM = 301 for all lactations in the learning set. For each observation time $(t)$ within a given lactation, a vector of forecast errors for all variables $(\boldsymbol{e_t})$ was generated. Simultaneously, a matrix describing the forecast co-variances $(\boldsymbol{Q_t})$ was generated, as described by West and Harrison (1997). Using the standard deviations $\left(\sqrt{Q_t^i}\right)$ derived from the forecast variances for each of the observed variables $(i)$, the values of the forecast errors for the individual variables $\left(e_t^i\right)$ where categorized in 1 of 4 observation categories $(Obs_i)$: low $\left(e_t^i < 0 - \sqrt{Q_t^i}\right)$, middle low $\left(0 > e_t^i > 0 - \sqrt{Q_t^i}\right)$, middle high $\left(0 < e_t^i < 0 + \sqrt{Q_t^i}\right)$, and high $\left(e_t^i > 0 + \sqrt{Q_t^i}\right)$.

The likelihoods for each of the possible observation categories within each variable, given the possible 2 states (mastitis positive (=clinical mastitis) or negative (= healthy)), was calculated as $p(Obs_i|Pos) = N_{Positive}^{Obs_i}/N_{Positive}$ and $p(Obs_i|Neg) = N_{Negative}^{Obs_i}/N_{Negative}$, where $N_{Positive}^{Obs_i}$ is the number of occurrences of a particular observation category when the observed cow is known to be mastitis positive, while $N_{Positive}$ is the total number of mastitis positive observations. Similarly, $N_{Negative}^{Obs_i}$ is the number occurrences of the observation category when the observed cow is known to be mastitis negative, while $N_{Negative}$ is the total number of mastitis negative observations.

For each of the 5 categorical variables, the likelihood of observing each possible category given the 2 mastitis conditions (positive or negative) were calculated in the same way as for the categorized forecast errors from the 7 continuous sensor variables.

## 2.5 Application of Naïve Bayesian Classification

For each milking, the 5 categorical and 7 continuous variables were combined using an NBC, and a posterior probability of the cow being mastitis positive was calculated according to Bayes' formula in equation 7.

$$p(Pos|Obs_1, \dots, Obs_i) = \frac{\sum_{i=1}^{12}\big(p(Obs_i|Pos)\big) \cdot p(Pos)}{\sum_{i=1}^{12}\big(p(Obs_i|Pos)\big) \cdot p(Pos) + \sum_{i=1}^{12}\big(p(Obs_i|Neg)\big) \cdot p(Neg)} \quad (7)$$

, where $p(Obs_i|Pos)$ is the probability of the observation of the $i^{th}$ categorized variable given that the cow is mastitis positive, $p(Obs_i|Neg)$ is the probability of the observation of the $i^{th}$ categorized variable given that the cow is mastitis negative, $p(Pos)$ is the prior probability that the cow is mastitis positive and $p(Neg)$ is the prior probability that the cow is mastitis negative. In the learning set, we observed that $p(Pos)$ = 0.5% and $p(Neg)$ = 99.5%.

If the observation of any variable was missing, then no likelihood related to that variable was included in the calculation of the posterior probability. The cow's milking was classified (predicted) as mastitis positive if the posterior probability was greater than a set threshold.

## 2.6 Performance Evaluation

All thresholds for positive classification between 0 and 1 with steps of 0.001 were evaluated. A posterior probability of mastitis above the set threshold was considered a mastitis alarm. For each threshold and milking, each mastitis alarm, or lack thereof, was compared to the diagnosis of mastitis provided by the farm staff.

In this study the morning and evening data were modeled separately and thus the performance of the DLM/NBC method was evaluated when applied separately to these 2 subsets. In this case, alarms were categorized as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) on a single milking basis, depending on the whether or not the farm staff had identified the cow as mastitis positive for that milking.

We also evaluated the performance achieved when combining the alarms produced by the 2 models. Because all mastitis cases had either been observed in the morning on the day they were registered or in the previous evening, we allowed for there to be 3 ways in which a TP would be assigned to a single mastitis observation: 1) an alarm was raised both by the morning model and the evening model, 2) an alarm was raised only by the morning model, 3) an alarm was raised only by the evening model. Any of these 3 scenarios would count as one TP observation. Similarly, 1 FN observation was counted if no alarm was raised by neither the morning model nor the evening model for a given mastitis observation. The FP and TN were still assigned on a single milking basis. These definitions are illustrated in Figure 2.

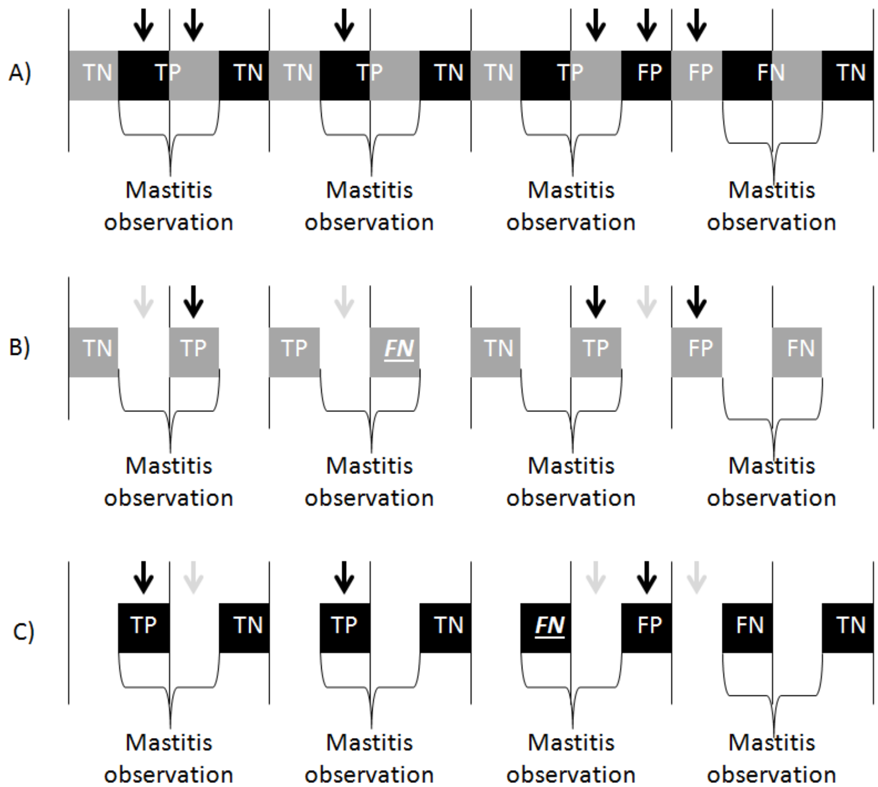



**Figure 2: Illustration of the definitions of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) alarms. Note that these examples only serve as illustrations and do not represent any actual observations. Grey blocks represent morning milkings, black blocks represent evening milkings, and the thin vertical lines represent the separations between days. Mastitis observations are always associated with a morning milking and the preceding evening milking. Black arrows represent alarms raised by the considered models (morning, evening, or both). Grey arrows represent alarms raised by the model, which is not considered (either morning or evening). A) When the alarms raised by the separate morning and evening models are both considered. B) When only alarms from the morning model are considered. C) When only alarms from the evening model are considered.**

This resulted in lists of sensitivities, calculated as $SE = TP/(TP + FN)$, and specificities, calculated as $SP = TN/(TN + FP)$. The lists of sensitivities and specificities, given the various thresholds, were used to plot receiver operating characteristics (**ROC**) curves (Zweig and Campbell, 1993), which was summarized by calculating the area under the curve (**AUC**) using the function auc from the library MESS in R.

In addition to the AUC of the ROC curve, the performances of the DLM/NBC method were evaluated based on the specificities achieved when sensitivity was set at 0.80, as recommended by Hogeveen et al. (2010), as well as the error rate of the predictions, calculated as $Error\ rate = (FP + FN)/(TP + FP + TN + FN)$. The 95 % confidence interval for sensitivity was calculated as $SE_{CI} = SE \pm \sqrt{(1 - SE)/(TP + FN)} \cdot 1.96$ and the 95 % confidence interval for specificity was calculated as $SP_{CI} = SP \pm \sqrt{(1 - SP)/(TN + FP)} \cdot 1.96$.

## 2.7 Sensitivity Analysis

The sensors used in this study come as 3 separate sensor packages, each of which can be obtained independently of the others. Package 1 included milk meters which measured milk yield and electrical conductivity. Package 2 included the AfiLabs which measured fat %, protein %, lactose %, blood %, and SCC category. Package 3 included the automatic scales which measured body weights. We measured the performance of the DLM/NBC method achieved by each of the 7 possible combinations of the 3 sensor packages, as well as the effect of omitting the always available 4 non-sensor variables, which we refer to as package 0. These 8 combinations of information packages are summarized in Table 2. Statistically significant differences between the specificities of 2 information packages were identified by comparing their 95 % confidence intervals.

**Table 2. The various combinations of the 4 information packages. Package 1: milk meter, package 2: AfiLab, package 3: automated scales, package 0: non-sensor information**

| Package combination | Included sensor packages | Included variables |
|---|---|---|
| 1-2-3-0 | Milk meter<br>AfiLab<br>automated scales<br>non-sensor information | Milk yield, conductivity, fat %, protein %, lactose %, blood %, SCC category, body weight, parity, mastitis history, season, week in milk |
| 1-2-3 | Milk meter<br>AfiLab<br>automated scales | Milk yield, conductivity, fat %, protein %, lactose %, blood %, SCC category, body weight |
| 1-2-0 | Milk meter<br>AfiLab,<br>non-sensor information | Milk yield, conductivity, fat %, protein %, lactose %, blood %, SCC category, parity, mastitis history, season, week in milk |
| 1-3-0 | Milk meter<br>automated scales<br>non-sensor information | Milk yield, conductivity, body weight, parity, mastitis history, season, week in milk |
| 2-3-0 | AfiLab<br>automated scales<br>non-sensor information | Fat %, protein %, lactose %, blood %, SCC category, body weight, parity, mastitis history, season, week in milk |
| 2-0 | AfiLab<br>non-sensor information | Fat %, protein %, lactose %, blood %, SCC category, parity, mastitis history, season, week in milk |
| 3-0 | Automated scales<br>non-sensor information | Body weight, parity, mastitis history, season, week in milk |

# 3 RESULTS

## 3.1 DLM-based Forecast Errors and Likelihoods

As is seen in Figure 3, milk conductivity, and to a lesser extend fat % and blood %, showed tendencies towards positive forecast errors, when mastitis occurred, while milk yield and lactose % showed tendencies towards being lower than forecasted by the DLM in mastitis positive cows. From a purely visual inspection of these plots, the presence of mastitis did not notably affect the tendencies of protein % and body weight towards being either above or below the forecasted value.

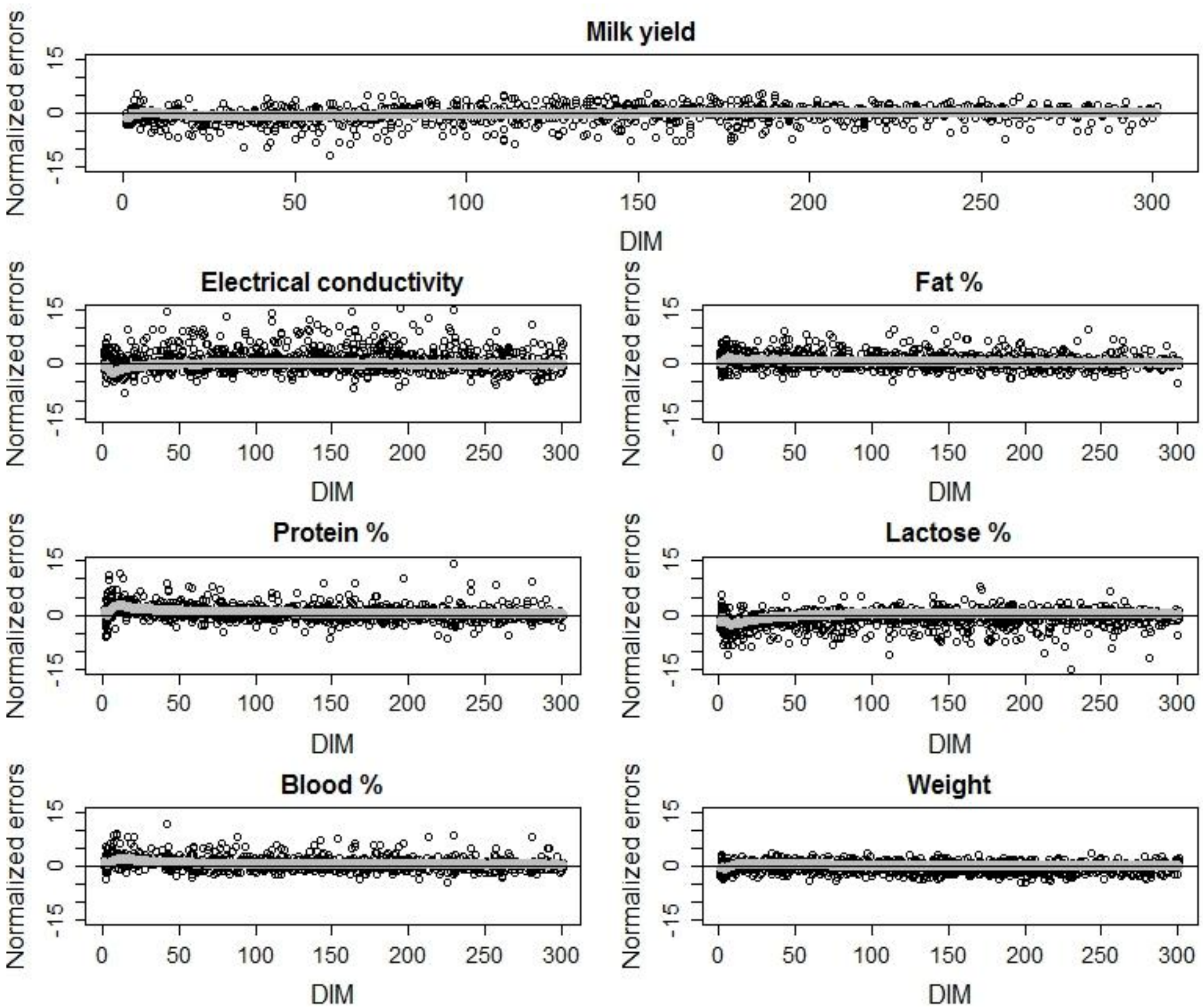



**Figure 3: Normalized forecast errors given health state for each of the 7 continuous sensors. Horizontal axis: days in milk (DIM). Vertical axis: normalized forecast errors. Points are errors when mastitis is observed. The thick grey lines are the average errors when no mastitis is observed. The thin horizontal line marks the position of 0.**

The likelihoods of the 5 categorical variables and the 7 categorized continuous variables are shown in Table 3. The likelihoods associated with both morning and evening milkings for the categorical variables were generally identical. For most continuous sensor data, the likelihoods differed considerably between morning and evening milkings.

**Table 3. Likelihood table for morning and evening milkings. p(Obs|Pos) is the likelihood of the observation given that the cow was diagnosed with mastitis. p(Obs|Neg) is the likelihood of the observation given that the cow was not diagnosed with mastitis**

| | | Morning | | Evening | |
|---|---|---|---|---|---|
| Variable | Observation | p(Obs\|Pos) | p(Obs\|Neg) | p(Obs\|Pos) | p(Obs\|Neg) |
| SCC[1] category | 0-200 | 0.50 | 0.82 | 0.50 | 0.82 |
| | 200-400 | 0.15 | 0.09 | 0.15 | 0.09 |
| | 400-800 | 0.11 | 0.04 | 0.11 | 0.04 |
| | 800+ | 0.24 | 0.06 | 0.24 | 0.06 |
| Previous Mastitis | No | 0.43 | 0.65 | 0.43 | 0.65 |
| | Yes | 0.57 | 0.35 | 0.57 | 0.35 |
| Parity | Later | 0.78 | 0.57 | 0.78 | 0.57 |
| | First | 0.22 | 0.43 | 0.22 | 0.43 |
| Season | Cold | 0.52 | 0.67 | 0.52 | 0.67 |
| | Warm | 0.48 | 0.33 | 0.48 | 0.33 |
| WIM[2] | 1 | 0.12 | 0.03 | 0.12 | 0.03 |
| | 2 | 0.04 | 0.03 | 0.04 | 0.03 |
| | 3+ | ~0.02 | ~0.03 | ~0.02 | ~0.02 |
| Milk yield | Low[3] | 0.31 | 0.09 | 0.31 | 0.21 |
| | Middle Low[4] | 0.37 | 0.59 | 0.32 | 0.43 |
| | Middle High[5] | 0.23 | 0.30 | 0.21 | 0.24 |
| | High[6] | 0.10 | 0.02 | 0.16 | 0.12 |
| Conductivity | Low[3] | 0.24 | 0.20 | 0.27 | 0.21 |
| | Middle Low[4] | 0.20 | 0.37 | 0.19 | 0.36 |
| | Middle High[5] | 0.17 | 0.32 | 0.17 | 0.29 |
| | High[6] | 0.40 | 0.11 | 0.37 | 0.14 |
| Fat % | Low[3] | 0.12 | 0.09 | 0.14 | 0.13 |
| | Middle Low[4] | 0.21 | 0.32 | 0.20 | 0.28 |
| | Middle High[5] | 0.28 | 0.40 | 0.23 | 0.35 |
| | High[6] | 0.39 | 0.19 | 0.43 | 0.25 |
| Protein % | Low[3] | 0.18 | 0.13 | 0.13 | 0.10 |
| | Middle Low[4] | 0.25 | 0.34 | 0.21 | 0.23 |
| | Middle High[5] | 0.28 | 0.37 | 0.29 | 0.37 |
| | High[6] | 0.30 | 0.16 | 0.37 | 0.29 |
| Lactose % | Low[3] | 0.38 | 0.16 | 0.69 | 0.59 |
| | Middle Low[4] | 0.29 | 0.36 | 0.17 | 0.27 |
| | Middle High[5] | 0.22 | 0.38 | 0.06 | 0.07 |

| | | | | | |
|---|---|---|---|---|---|
| | High[6] | 0.11 | 0.11 | 0.07 | 0.07 |
| Blood % | Low[3] | 0.06 | 0.06 | 0.18 | 0.21 |
| | Middle Low[4] | 0.20 | 0.27 | 0.18 | 0.20 |
| | Middle High[5] | 0.38 | 0.50 | 0.24 | 0.27 |
| | High[6] | 0.36 | 0.18 | 0.40 | 0.32 |
| Body weight | Low[3] | 0.47 | 0.31 | 0.29 | 0.15 |
| | Middle Low[4] | 0.26 | 0.31 | 0.22 | 0.22 |
| | Middle High[5] | 0.18 | 0.25 | 0.24 | 0.28 |
| | High[6] | 0.09 | 0.13 | 0.26 | 0.36 |

[1]SCC is listed as thousands of cells per ml.
[2] Week in milk, from 1 to 43. After week 3, the likelihood given mastitis condition is approximately constant.
[3]The observed value is more than one standard deviation below the forecasted value
[4]The observed value is less than one standard deviation below the forecasted value
[5]The observed value is less than one standard deviation above the forecasted value
[6]The observed value is more than one standard deviation above the forecasted value

### 3.2 Posterior Probabilities

Figure 4A shows the average posterior probabilities produced with the DLM/NBC method for lactations with no mastitis cases based on all available sensor and non-sensor data (Table 2, Package combination 1-2-3-0). Morning and evening milking probabilities are shown sequentially. In the beginning of the lactation, the DLM-informed probability of mastitis is fluctuating and generally high, before it settles to a more constant level of nearly 0. This is a typical pattern, as the DLM is adapting to the data of the individual cow during these first few days. Thus there is an elevated risk of FP in this stage of the lactation, which on average takes 15 d, as indicated by the vertical line.

Figure 4B shows a lactation where a mastitis events is observed, marked by the dashed vertical line on DIM = 103. Notice the sharp spikes in the mastitis probability coinciding with this event, followed by at least 14 days where the posterior probability is not considered. Notice that spikes in the probability start to occur a few milkings before the event is observed. In this study, those premature spikes would be counted as FP, but they could conceivably be used for early warnings. Also, similarly to Figure 4A, fluctuations and higher-than-normal mastitis probabilities are seen at the beginning of the lactation, in spite of there being no observed mastitis event.

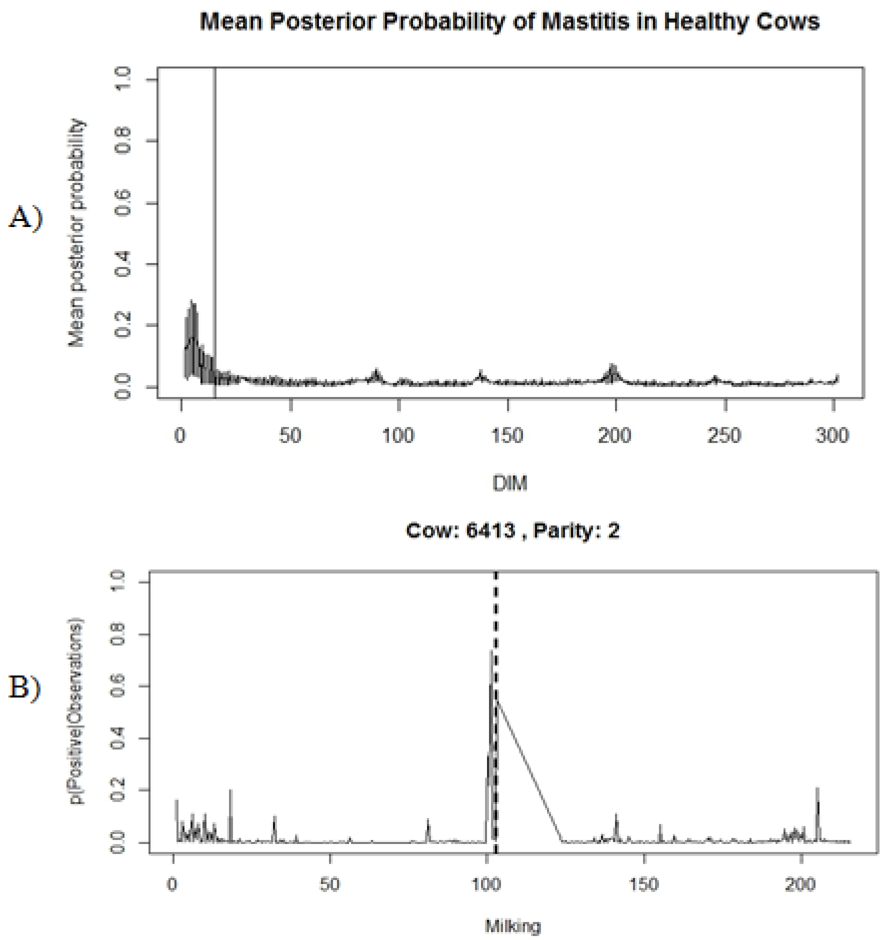



**Figure 4: A) The average posterior probability of mastitis given DIM for healthy lactations. The vertical line indicates DIM = 15. B) Posterior probabilities of mastitis for one cow during a single lactation. The thick, dashed vertical line on DIM = 103 indicates the observation of a mastitis event. Probabilities for morning and evening observations are plotted sequentially in both plots.**

### 3.3 Predictive Performance

Table 4 shows the predictive performances of the 8 information packages described in Table 2. Table 4 is sorted by the error rate, specificity, and AUC, which were achieved by combining the alarms of the morning and evening models. In all cases, sensitivity was held at 0.80 with the 95% confidence intervals being from 0.77 to 0.83 for morning, evening, and the combined alarms. The best performance of the DLM/NBC method using the full set of available sensor and non-sensor information (combination 1-2-3-0 in table 2) resulted in an AUC of 0.89 with a specificity of 0.81 and an error rate of 0.19.

**Table 4: The predictive performance of the 8 combinations of information packages as measured by the area under the receiver operating characteristic curve (AUC), specificity (SP) and error rate. Sensitivity was kept at 0.80. Package 1 includes the milk meter, 2 includes the AfiLab, 3 includes the automated scale, and 0 includes of non-senor information**

| Package Combination | Milking Time of Day | AUC | SP | SP C.I. 95% | Error rate |
|---|---|---|---|---|---|
| 1-2-3-0 | Combined | 0.89 | 0.81 | 0.809-0.811 | 0.19 |
| | Morning | 0.85 | 0.75 | 0.748-0.752 | 0.25 |
| | Evening | 0.73 | 0.50 | 0.498-0.502 | 0.51 |
| 1-2-0 | Combined | 0.89 | 0.81 | 0.809-0.811 | 0.19 |
| | Morning | 0.85 | 0.75 | 0.748-0.752 | 0.26 |
| | Evening | 0.73 | 0.50 | 0.498-0.502 | 0.49 |
| 1-2-3 | Combined | 0.88 | 0.79 | 0.789-0.791 | 0.21 |
| | Morning | 0.82 | 0.68 | 0.678-0.682 | 0.33 |
| | Evening | 0.67 | 0.42 | 0.418-0.422 | 0.60 |
| 2-3-0 | Combined | 0.85 | 0.75 | 0.749-0.751 | 0.25 |
| | Morning | 0.79 | 0.64 | 0.638-0.642 | 0.35 |
| | Evening | 0.75 | 0.55 | 0.548-0.552 | 0.47 |
| 2-0 | Combined | 0.85 | 0.74 | 0.739-0.741 | 0.25 |
| | Morning | 0.78 | 0.63 | 0.628-0.632 | 0.37 |
| | Evening | 0.76 | 0.56 | 0.558-0.562 | 0.46 |
| 1-3-0 | Combined | 0.86 | 0.74 | 0.739-0.741 | 0.25 |
| | Morning | 0.82 | 0.65 | 0.648-0.652 | 0.34 |
| | Evening | 0.70 | 0.49 | 0.488-0.492 | 0.52 |
| 1-0 | Combined | 0.86 | 0.73 | 0.729-0.731 | 0.27 |
| | Morning | 0.81 | 0.63 | 0.628-0.632 | 0.38 |
| | Evening | 0.74 | 0.53 | 0.528-0.532 | 0.47 |
| 3-0 | Combined | 0.76 | 0.58 | 0.579-0.581 | 0.42 |
| | Morning | 0.73 | 0.53 | 0.528-0.532 | 0.46 |
| | Evening | 0.70 | 0.50 | 0.498-0.502 | 0.50 |

The combination of morning and evening alarms consistently outperformed alarms raised based on either model alone. This was true for all 3 performance measures. Similarly, the 3 performance measures consistently showed that the morning model alone outperformed the evening model alone. These relationships are also evident from the ROC curves in Figure 5A.

From the ROC curves in Figure 5B we see that the predictive performances achieved with the 8 information packages fall into 3 distinct groupings. The highest ROC curves are for the combinations 1-2-3-0, 1-2-0, and 1-2-3. In terms of all 3 performance measures, combinations 1-2-3-0 and 1-2-0 yield exactly the same performances for the combined alarms, while the specificity of combination 1-2-3 is significantly lower the other 2 at the 95% confidence level.

A)

B)

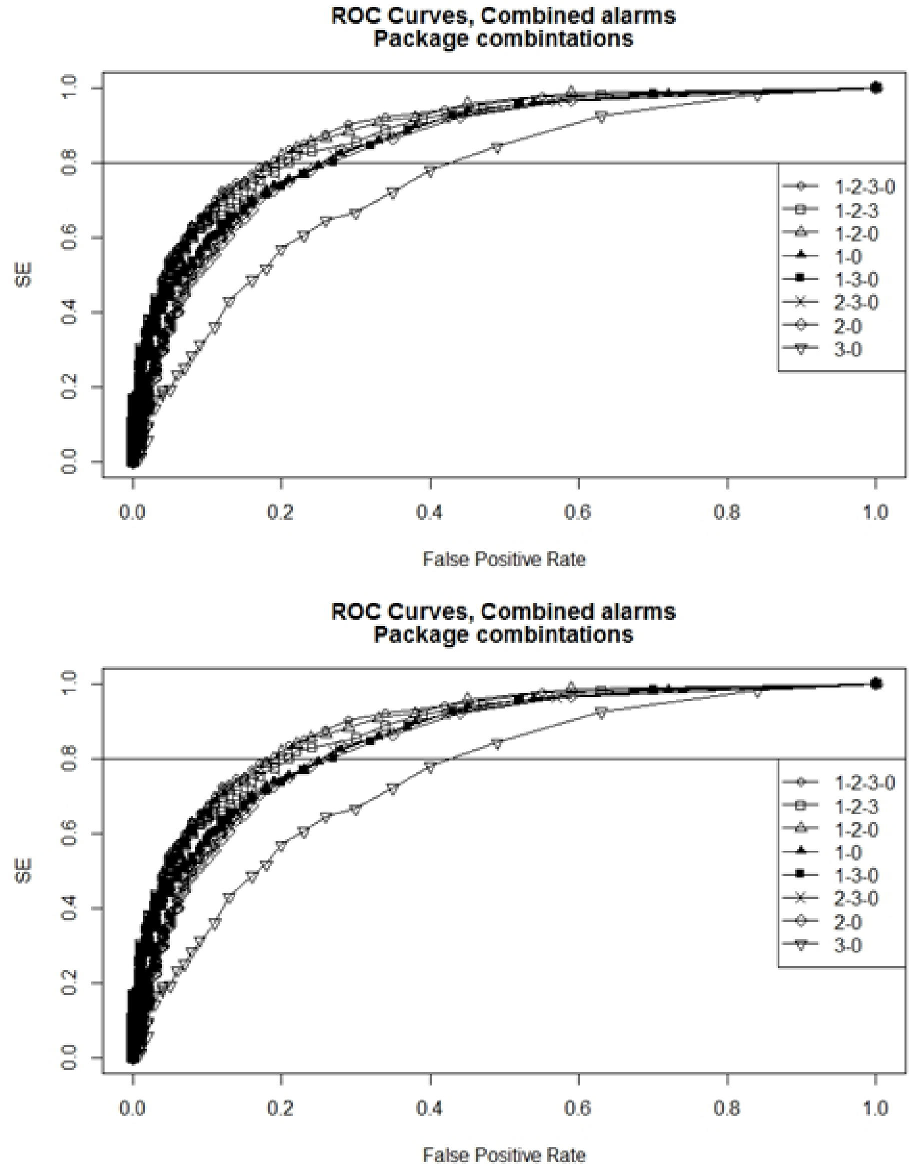



**Figure 5: Receiver operating curves (ROC) for mastitis prediction. A) ROC curves for the combined alarms, the morning alarms and evening alarms, when all sensor and non-sensor information is included. B) The ROC curves for the combined alarms, given the 8 information package combinations included in this study. Package 1 includes the milk meter, package 2 includes the AfiLab, package 3 includes of the automated scale, and package 0 includes of non-sensor information.**

The second-highest group of ROC curves consists of the combinations 2-3-0, 1-3-0, 2-0, 1-0. When sensitivity is held at 0.80, the specificities and error rates of combinations 3-1-0 and 2-0 were similar, although the AUC was slightly greater for the 3-1-0 combination (difference 0.01). Even though the AUC of the 1-0 combination was greater than that of the 2-3-0 combination, the 2-3-0 combination achieved preferable specificity and error rate values, when the sensitivity was set at 0.80.

The third group consists solely of combination 3-0, showing the least favorable values with respect to all 3 performance measures.

## 4 DISCUSSION

Our main objective was to describe and demonstrate the combination of a multivariate DLM and an NBC as a novel but intuitive method for combining sensor and non-sensor data for detecting mastitis. We further assessed the performance which can be expected from this method when some sensor packages are not available.

### 4.1 DLM/NBC Methodology

The basic premise behind the NBC is that all included variables are mutually independent and are only affected by the outcome of interest, but e.g. milk yield, the various milk components and conductivity are known to be highly correlated (Yoshida et al., 2005). The interdependencies between the 7 continuous variables are however accounted for by the act of co-modeling these variables in the multivariate DLM, taking into account the observational and systematic co-variances between them, as described by the $V$ and $W$ matrices, respectively. No interdependencies between the 5 categorical variables are taken into account in our DLM/NBC method. For example, it is known that season as well as parity affects milk composition (Yang et al., 2013). Never the less, the NBC can yield excellent performances, even if the assumption of independence is violated (Pazzani, 1997). Furthermore, Steeneveld et al. (2010) attempted to improve their classification of TP mastitis cases by expanding their naïve Bayesian network to include dependencies between their included variables, but found that the resulting classification performance was not improved. We thus consider the multivariate DLM/NBC method to be a reasonable compromise between accounting for dependencies between continuous variables while still allowing for easy incorporation of all available data, including the categorical non-sensor variables.

### 4.2 Performance of the DLM/NBC Method and Contribution from Sensor Packages

As seen in table 4, the omission of the non-sensor information leads to a reduction in all 3 measures of predictive performance. This is in agreement with the finding of Steeneveld et al. (2010) that including non-sensor information to distinguish between TP and FP mastitis alarms, raised by an automatic milking system, reduced the number of false positives by 35 %, thus increasing the specificity. Furthermore, while Steeneveld et al. (2010) included similar non-sensor data as were included in this study, they also included several other sources of non-sensor information, which were not available for this study, and which they found to have considerably higher significance in distinguishing between TP and FP mastitis alarms. Thus it stands to reason that even better performances could be achieved by including still more information.

Package combination 3-2-1-0 yielded the same performance measures as package combination 2-1-0, which suggests that if a farmer uses milk meters and the AfiLab, there is nothing more (in terms of mastitis detection) to be gained from investing in the automated scales. This does not mean that the body weight contains no information, however. This is evident from the fact that package combination 2-3-0 performs better than 2-0 in terms of specificity and package combination 1-3-0 perform better than 1-0 in terms of specificity as well as error rate. As for package 1 compared to package 2, it is worth noticing that the AUC is higher for package combination 1-0 compared to combination 2-0 (AUC = 0.860 vs. 0.848). One might thus be inclined to think that the milk meter provides better mastitis detection than the AfiLab, if one had to choose to have only one of these sensors. This is however not the case when sensitivity is held at 0.80, as the corresponding specificity and error rate are both favorable for the AfiLab, with the difference in specificity being outside the 95% confidence interval.

A clear cost-benefit analysis of investment in sensor packages was not feasible because real market prices were not available to us and sensors are usually also used for other purposes than mastitis detection. Routine maintenance time and costs were negligible for all sensors.

### 4.3 Results of the DLM/NBC Method Compared to the Literature

A direct comparison between our results and those reported in other studies does not fairly judge the performance of various methods, given that the different studies include data of different origins and unknowable quality. For example, case definitions of mastitis are not standardized and time windows of detection vary (Hogeveen et al., 2010). We do however consider the following considerations to be worthwhile for comparison.

Some authors reported greater mastitis specificities than we found. For example, de Mol et al. (2001), Kamphuis et al. (2010) and Mollenhorst et al. (2010) presented mastitis detection with specificities of 0.979, 0.987 and 0.990, respectively. The associated sensitivities were 0.67, 0.32 and 0.474, respectively, as compared to our set sensitivity of 0.80. Greater specificity results in lower sensitivity. Figure 5 shows that higher specificities are obtainable at the cost of lower sensitives.

Other authors presented both high sensitivities and specificities, but these authors typically used wider time windows in their performance evaluation, as opposed to the 2 milkings window used in our study when we combined the morning and evening alarms. For example, de Mol et al., (1997) used a time window of 17 d around the day of diagnosis (-10 d to +7 d, sensitivity 0.90, specificity 0.982), while Cavero et al. (2009) used a 5-d window (-2 to +2, sensitivity 0.929, specificity 0.939). Such long time windows may not be useful in practice. If an alarm is raised several days before clinical signs of mastitis are visible, the farmer will likely believe the alarm to be a FP. If this happens too frequently, the farmer will lose trust in the system, which would make it worthless (Hogeveen et al., 2010). Conversely, if an alarm is not raised until several days after clinical signs appear, treatment will be needlessly delayed. Thus a very narrow time window is needed for practical mastitis detection applications.

We found only 3 studies with a combination of predictive performance and narrow time windows comparable to those presented in this paper. These studies presented mastitis detection based on artificial neural networks (Nielen et al., 1995; Sun et al., 2010) and fuzzy logic (Kamphuis et al., 2008). These 3 studies were all conducted with relatively small study populations, which invariably reduces the reliability of any results. Specifically, Nielen et al. (1995) included only 55 cows (31 with mastitis) and Kamphuis et al. (2008) included 18 mastitic cows. Given the reported number of cases and non-cases in these studies, the 95 % confidence intervals for the reported sensitivities would include values as low as 0.71 and 0.59 for Nielen et al. (1995) and Kamphuis et al. (2008), respectively. Sun et al. (2010) considered 194 cows, of which 43 (88 udder quarter milkings) were actually observed to be mastitic. However, Sun et al. (2010) inflated the number of infected quarter milkings to 895 by assuming that the quarters of the 43 cows were also mastitic at other times than indicated in the log book, if the observed milk yield, conductivity and SCC showed values above or below specific thresholds. They subsequently trained neural networks to detect mastitis, based on observed values of milk yield and conductivity, thus making the detection depend on the same variables that were used to define the majority of the events. Thus, the impressive performance (SE = 0.87, SP = 0.91) shown by Sun et al. (2010) is likely due to incorporation bias.

A common approach in studies like the ones cited above is to clear the data set of missing data before applying a detection method. This will invariably give an unfairly favorable impression of the performance which ultimately cannot be transferred to practice where missing data are unavoidable. In contrast, the DLM/NBC method is cable of handling missing observations, as was demonstrated on our realistic data sets with all instances of missing data preserved.

### 4.4 Perspectives

The performance of the DLM/NBC method might be improved if milk yield was corrected for the interval between milkings. Longer intervals are associated with greater milk yield (Palmer et al., 1994). The effects of other design choices, such as the selection of “healthy” lactations to estimate variance components, should also be further evaluated. The expectation maximization algorithm is time consuming, which can be a hindrance for the practical application of the DLM/NBC method. In this study, we chose to only include high quality data when estimating the variance components, in part to reduce computation time. It would make sense to study the effect of the amount versus the quality of the data with respect to final model performance, which to our knowledge has not been done elsewhere. An alternative method could be to assume an unknown, non-constant system variance, which would be continuously estimated as described by West and Harrison (1997).

Given the performance demonstrated in this study, it is reasonable to investigate the method’s value for detecting other conditions in dairy cows, or indeed health states in other production animals, such as pigs or poultry. Furthermore, one of the main advantages of using the NBC for information integration was the relative simplicity with which specific data could be ignored (for example in the case of missing data) or added to the probability calculation. A relevant follow-up study would be to investigate how long after the inclusion of a new sensor the information collected by that sensor would add significantly to the performance of the DLM/NBC method, if the likelihoods need to be learned from observations in that herd. If the likelihoods of a condition, associated with the values of the sensor, can be directly applied between different herds, then a new sensor could potentially be useful immediately after its integration. However, if the likelihoods have to be estimated using on-site data, then the time before relevant information is added will depend heavily on the prevalence of the condition of interest.

## 5 CONCLUSIONS

In this study we showed that a combination of a multivariate DLM to produce forecasts and an NBC using mastitis-dependent likelihoods of forecast errors can be meaningfully used to combine multiple types of data for detecting mastitis in dairy cows. An advantage of the proposed method is the ease with which missing observations can be handled, and information from new sensors added, which is a necessary ability in real world farm situations. With this DLM/NBC method, we reached an AUC of 0.89, with a specificity of 0.81 when the sensitivity was held at 0.80 and when using the alarms raised during both morning and evening milkings.

We tested the predictive performance using all combinations of the 3 available sensor packages as well as the omission and inclusion of non-sensor data. While all sensor packages held some information relevant for mastitis detection, the automated scale was by far the least informative. Including non-sensor data significantly improved the performance.

## 6 ACKNOWLEDGMENTS

We thank the staff of the University of Florida Dairy Unit for their help with collecting the data for this study, in particular Eric Diepersloot and Grady Byers. Dan Jensen and Albert De Vries thank Henk Hogeveen for his hospitality during our common stay at Wageningen University in November 2014. Dan Jensen was financially supported from The Danish Council for Strategic Research (The PigIT project, Grant number 11-116191). Albert De Vries received financial support from USDA-NIFA AFRI grant award number 2013-68004-20424 entitled "Southeast Quality Milk Initiative: Controlling mastitis and Improving Milk Quality". We also thank veterinarian Line Svenneson for sharing her knowledge of the practical aspects of information availability on dairy farms.

## 7 REFERENCES

Cavero, D. et al., 2009. Mastitis and lameness detection in dairy cows by application of fuzzy logic. *Livestock Science*, 125, pp.92–96.

Geenen, P.L. et al., 2011. Constructing naive Bayesian classifiers for veterinary medicine: A case study in the clinical diagnosis of classical swine fever. *Research in Veterinary Science*, 91(1), pp.64–70. Available at: http://dx.doi.org/10.1016/j.rvsc.2010.08.006.

Gromiha, M.M. & Suresh, M.X., 2007. Discrimination of mesophilic and thermophilic proteins using machine learning algorithms. , pp.1274–1279.

Hogeveen, H. et al., 2010. Sensors and clinical mastitis--the quest for the perfect alert. *Sensors (Basel, Switzerland)*, 10(9), pp.7991–8009. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3231225&tool=pmcentrez&rendertype=abstract [Accessed January 22, 2015].

Huybrechts, T. et al., 2014. Early warnings from automatic milk yield monitoring with online synergistic control. *Journal of dairy science*, 97(6), pp.3371–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24731631.

Jensen, D.B. et al., 2012. Bayesian prediction of bacterial growth temperature range based on genome sequences. *BMC genomics*, 13 Suppl 7(Suppl 7), p.S3. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3521210&tool=pmcentrez&rendertype=abstract [Accessed January 17, 2013].

Kamphuis, C. et al., 2008. Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. *Journal of dairy science*, 91(12), pp.4560–4570. Available at: http://dx.doi.org/10.3168/jds.2008-1160.

Kamphuis, C. et al., 2010. Decision-tree induction to detect clinical mastitis with automatic milking. *Computers and Electronics in Agriculture*, 70, pp.60–68.

Kaniyamattam, K. & De Vries, a, 2014. Agreement between milk fat, protein, and lactose observations collected from the Dairy Herd Improvement Association (DHIA) and a real-time milk analyzer. *Journal of Dairy Science*, 97(5), pp.2896–2908. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24630652.

De Mol, R.M. et al., 2001. Detection of estrus and mastitis: Field performance of a model. *Applied Engineering in Agriculture*, 17(3), pp.399–407.

De Mol, R.M. et al., 1997. Results of a multivariate approach to automated oestrus and mastitis detection. *Livestock Production Science*, 48, pp.219–227.

Mollenhorst, H., van der Tol, P.P.J. & Hogeveen, H., 2010. Somatic cell count assessment at the quarter or cow milking level. *Journal of dairy science*, 93(7), pp.3358–3364. Available at: http://dx.doi.org/10.3168/jds.2009-2842.

Nielen, M. et al., 1995. Comparison of analysis techniques for on-line detection of clinical mastitis. *Journal of dairy science*, 78, pp.1050–1061.

Ostersen, T., Cornou, C. & Kristensen, A.R., 2010. Detecting oestrus by monitoring sows' visits to a boar. *Computers and Electronics in Agriculture*, 74(1), pp.51–58. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0168169910001201 [Accessed April 12, 2013].

Pazzani, P.D.M., 1997. Beyond independence: Conditions for the optimality of the simple Bayesian classifier. *Machine Learning*, 29, pp.103–130.

Sun, Z., Samarasinghe, S. & Jago, J., 2010. Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks. *The Journal of dairy research*, 77, pp.168–175.

The R Core Team, 2013. *R : A Language and Environment for Statistical Computing* 2.15.3 ed., R Foundation for Statistical Computing.

West, M. & Harrison, J., 1997. *Bayesian Forecasting and Dynamic Models* 2nd ed., New York, USA: Springer.

Yang, L. et al., 2013. Effects of seasonal change and parity on raw milk composition and related indices in Chinese Holstein cows in northern China. *Journal of dairy science*, 96, pp.6863–9.

Yoshida, T., , Lopez-Villalobos, N.. & Holmes, C. w., 2005. Relationships between milk yield and fertility in dairy cattle. *Journal of dairy science*, 65, pp.143–146.

Zweig, M.H. & Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4), pp.561–577.

# Chapter 10: Discussion and Perspectives

The objectives of this PhD project introduced a number of challenges, as presented section 1.2. Here the chosen solutions to those challenges, along with the findings achieved with those solutions, are discussed in relation to the existing scientific literature. These discussions raise perspectives for future research, which will be suggested during the discussion as well as summarized at the end of this chapter.

## 10.1 Combining diverse data streams

As stated in section 1.2, a major challenge raised by the goals of this PhD project, was the matter of how to combine the very diverse data streams, which were collected at the farms. From the commercial Danish pig farm, these data streams included continuous animal-oriented and environment-oriented sensor data, while the data streams from the University of Florida dairy herd included continuous animal-oriented sensor data as well as categorical sensor and non-sensor data.

For continuous time series data, such as the sensor data used in the papers presented in this thesis, it is often a meaningful pre-processing method to have some model learn the expected pattern of the data stream, and then for each observation compare the forecasted values with the observed values. In Papers 3, 4, and 5, this was achieved with multivariate DLM, while the time series element was simply ignored in Paper 2. The multivariate DLM is a relatively modern and not yet widely used method for handling time series data. More classical time series handling methods include moving averages, as used by *e.g.* Cavero et al. (2008) and Kamphuis et al. (2010). Compared to these, the multivariate DLM present a number of advantages. For one thing, the multivariate DLM will by design include some estimate of the co-variances between the various observable and unobservable parameters included in the model.

These co-variances can be estimated in different ways. The simplest solution is to assume no co-variances between the various parameters, as was done by *e.g.* Madsen and Kristensen (2005). Another option is to estimate the variance-covariance matrices using the expectation maximization (EM) algorithm, as described by West and Harrison (1997). This was the solution selected in Papers 3, 4, and 5. As mentioned in section 4.3, this inclusion of co-variances enables the model to take the interconnectedness between the modeled variables into consideration when making new forecasts. This also means that even when one or more observations of one variable is missing, the model is still able to predict changes in the level and the evolution of that variable, based on the changes in those variables which are observed.

Although the use of DLMs in animal science, including PLF, is not yet widely seen, the fact that information from the sensor data could be extracted via dynamic modeling is not in itself a novel discovery, as this has previously been shown to be useful for detection of relevant events in dairy cows (de Mol et al. 2001; de Mol et al. 1997), sows (Ostersen et al. 2010; Cornou & Lundbye-Christensen 2012), and that they are probably useful for detecting health problems in weaned pigs by monitoring their drinking behavior (Madsen et al. 2005). The novel part compared to these studies, however, is the ways in which the forecast errors were further processed in order to raise alarms concerning the undesired events, which will be discussed further in a later section of this chapter. Another major insight, which was first gained from the work related to Paper 3 and does not seem to have been considered in the above listed studies, was that the variances of the data in the separate data streams needed to be adjusted to approximately the same level. Not doing this meant that the EM algorithm would be unable to estimate the variance-covariance matrices, because certain values would become computationally singular. A workable solution was to manually upscale or downscale the data in the individual data streams by factors of either 10 or 100 to bring the overall variances within the separate data streams to approximately the same level. This method was subsequently used in Papers 4 and 5 as well.

Furthermore, the forecast errors which are produced by the DLM can be further processed in several ways before being parsed to a classification model. As an example, de Mol et al. (1997) used simple thresholds based on various confidence intervals, where just one of the forecast errors had to exceed the threshold for an alarm to be raised. In Papers 3 and 4, the forecast errors were unified via Cholesky decomposition and alarms were raised based on a threshold value for this unified error. And in Paper 5, the forecast errors were categorized based on whether they were positive or negative, and whether they deviated from zero by more or less than one standard deviation, also known as 1σ, corresponding to a ~68 % confidence level. The classification cutoffs could also have been based on *e.g.* 2σ or 3σ (~95 % and ~99.7 % confidence level, respectively), or more categories could have been made by using all of these cutoffs. Alternatively, the forecast errors could have been used in their numerical form rather than categorizing them, or the cumulative sum of the forecast errors could have been considered instead. Testing these various forecast error handlings in combination with different classification tools, with the purpose of optimizing the predictive performance, would be interesting goals for future research.

#### 10.1.1 Animal-oriented vs. environment-oriented sensor data

When taken together, the findings in Papers 2, 4, and 5 give the distinct impression that animal-oriented and environment-oriented data are best pre-processed in different ways. Specifically, the information from the behavior-related data such as the drinking behavior of the pigs in Paper 4, as well as the physiological data such as the milk yield and EC of the cows in Paper 5, could all be extracted meaningfully by means of dynamically modeling the data followed by parsing the forecast errors produced by the dynamic models to some classification model. In Paper 2, a strategy of daily temperature summary statistics combined through a logistic regression model was seen to be an effective method of extracting information related to the undesired events. This is in stark contrast to what was seen in Paper 4, where the DLM/Cholesky method applied to environmental variables such as temperature was found to be little more than useless.

At this point, a recap of the main differences between the DLM and the summary values is in order: for each hour, where the temperature is observed, the DLM will make a forecast of what value is expected. This forecast is based on the values observed up until that point in time and the functions describing the evolution of the temperature, which in the case of Paper 4 was a single harmonic wave. In other words, the DLM has no choice but to consider the individual observations in relation to a wider context. This is important because the forecast errors, *i.e.* the differences between what is observed and what was forecasted, are the basis for raising the alarms. When making the summary values, as was done in Paper 2, we look at absolute values, disregarding previous observations and any other context. As an example, assume that a 4 $^{o}$C increase in temperature is seen from one hour to the next. If the DLM for whatever reason has come to expect a temperature increase of approximately 4 $^{o}$C, this event would be considered normal and no alarms would be raised. To the pigs, however, this increase in temperature is still likely to cause a stress response (Lopez et al. 1991a), which in turn might manifest as an undesired event such as the onset of diarrhea (Shimizu et al. 1978), regardless of what wider context the change happened to fit into. In other words, if the pigs are likely to experience the different variables in absolute rather than relative terms, then we too, with our models and alarm systems, should consider those variables in absolute terms.

As described in section 1.3.2, environmental variables such as temperature and humidity are generally monitored for the purpose of climate control and not directly for the purpose of detecting or providing early warnings about undesired events. Probably for this reason, no scientific literature, which was concerned with how the environment should be monitored for this purpose, could be found. In one study concerned with monitoring the ear skin temperature, the skin temperature was modeled using a linear mixed-effects model, but for descriptive and not predictive purposes (Andersen et al. 2008). It would thus seem that this insight, that environmental variables are better considered with summary statistics rather than deviations from model-based expectations, is genuinely new. As such, it should of course be verified rigorously to eliminate alternative explanations.

One such alternative explanation could be that the forecast errors from an environment-oriented DLM need to be considered with different prediction windows compared to animal-oriented DLM forecast errors in order to exploit their full potential. Under the environment hypothesis, as defined in section 1.4, the environmental variables are (or can be) the causing factors for undesired events such as diarrhea or pen fouling. Under the normality hypothesis, likewise defined in section 1.4, changes in behavior and physiology are expected to precede the onset of visible problems, including those that are caused by the environment. In other words, changes in the environment would be expected to precede changes in behavior and physiology, which in turn are expected to precede the directly observable physical problems.

Another alternative explanation, which would have to be tested, is that the parts of the DLM related to the environmental variables were not optimally defined for the purpose of detecting undesired events. For example, it could be that the model is actually too accurate; from Paper 4 Figure 4 it is seen that the temperatures on average follow season-specific harmonic waves in a diurnal pattern, and the model was designed to reflect this. It is known from the literature, however, that pigs are best served with constant temperatures (Lopez et al. 1991b; Lopez et al. 1991a; Shimizu et al. 1978), and thus a DLM which reflect this fact may provide more informative forecast errors. In other words, when it comes to modeling the environment-oriented data, it may be more useful to model the environment which is preferable to the pigs rather than the environment which is actually expected. If this is the case, however, it might be difficult to argue for how exactly that would differ from considering simple summary statistics.

### 10.1.2 Categorical data

For Paper 5, the categorical sensor and non-sensor data from the University of Florida Dairy herd were combined with the forecast errors for the continuous data by means of a naive Bayesian classifier (NBC). Thus the DLM/NBC method served as a method for combining animal oriented sensor data, which can be meaningfully modeled with a DLM, with other data for which the DLM-approach is apparently not very useful. Remember that this was found to be a major apparent issue in Paper 4. As an example, the somatic cell count (SCC) data is a good analogy to the temperature data from Paper 4, even though SCC is animal-oriented sensor data and temperature is environment-oriented sensor data. In both cases we are dealing with data collected over time, meaning that they could in principle be modeled using a DLM. In both cases, however, it has been found more useful to summarize the observations. In the case of temperature, this was done by finding the absolute maximum and minimum temperatures, as well as the fastest increases and decreases in temperature over the period of one day. Thus the semi-continuous measurements are summarized by four numeric values per day per temperature sensor. For the SCC, one measurement was made per milking, but due to great uncertainties about these values, they were summarized by being placed into one of four predefined categories.

NBCs have been used in PLF contexts before, such as to improve the sensitivities of the mastitis alarms raised by existing  automatic milking systems (Steeneveld et al. 2010). In this sense, the use an NBC was not overly novel. The novel part was to show that the information extracted from the sensor data (via the DLM) could be directly combined with the information from the categorical data. In this way, the DLM/NBC method skips a classification step compared to the one presented by Steeneveld et al. (2010). Which of these strategies is best is however hard to tell, given the rather different nature of the approaches, combined with the inherent uncertainty about the assigned gold standards for true positives and true negatives in the two different datasets.

## 10.2 Comparison of the various classification models

As explained in section 1.2, the classification models were those tools which determined if an alarm should be raised or not, based on the pre-processed data. A total of three such classification tools were attempted for detection and/or early warning of undesired events, namely logistic regression, forecast error unification through Cholesky decomposition and transformation, and NBC. This section will discuss the potentials and limitations of these methods.

### 10.2.1 Logistic regression

The summary/logistic regression method was shown in Paper 2 to be useful for extracting information from the environment oriented sensor data. Specifically, the part of this method which makes it useful for this purpose is the summary part, as discussed in section 10.1.1. The logistic regression simply served as a way of combining the summery data in such a way that a single posterior probability of undesired events could be calculated. In this sense, the logistic regression served the same principle purpose as the NBC from the DLM/NBC method in Paper 5, and those two classification tools are arguably closely related. This also means that *e.g.* an NBC or some other classification tool could have taken the place of the logistic regression in Paper 2. The logistic regression method is characterized, and limited, by being a linear classifier. As was seen from section 1.3.3, there are plenty of other non-linear classification tools to choose from, such as NBCs, artificial neural networks or decision trees, as well as methods which were not mentioned in that sections, such as support vector machines. It is conceivable that combining one of these tools with a relevant combination of environment-oriented summary statistics could yield even better performances of detection and/or early warning of the undesired events, compared to using the logistic regression. This assertion is in line with the findings of Nielen et al. (1995a), who found that a logistic regression model and an ANN achieved similar performances when trained on the same data, but that the specificities were generally higher for the ANN.

Testing different classification tools in combination with summary statistics for environment-oriented data would be a relevant objective for future studies. Such future studies should include summary values of the (section level) humidity, as well as take into account the fact that the optimal temperature and humidity will change as the pigs grows, which was not considered in Paper 2. These (probably) improved performances should be compared to those achieved when attempting to make better environment-oriented DLMs as part of the efforts to rigorously test the apparent finding that a DLM is an inferior pre-processing method for environment-oriented data compared to summary statistics, as discussed in detail in section 10.1.1.

Finally, it is worth mentioning the fact that the summary/logistic regression method was implemented in such a way that it could not distinguish between diarrhea and pen fouling. Nevertheless, having this option of discriminating between different events could be achieved simply by training the logistic regression to recognize each of the events of interest.

#### 10.2.2 Forecast error unification by Cholesky decomposition

One important limitation of the DLM/Cholesky method used in Papers 3 and 4 is that it cannot even in principle distinguish between different undesired events, unlike other methods such as the logistic regression method discussed above. As explained in section 3.4, this is because the DLM/Cholesky method will base its alarms on how different the observed system is from the normal system, but does not distinguish between which parts of the system (*i.e.* which sensors) show deviations from normalcy or in which direction (positive or negative) the change is leaning. This can however be both a disadvantage and an advantage, depending on what type of alarms a farmer desires and how the method is implemented. If, for example, the farmer wishes to get specific alarms concerning diarrhea, pen fouling, tail biting, or any number of other specific problems he might face, then the DLM/Cholesky method will not do him much good. On the other hand, if the farmer does not care about event specific alarms but just want to know where to be extra aware of signs of any and all problems, a DLM/Cholesky method, made to include all available (animal oriented) data, would be able to provide just this. It might also be that the farmer could use a combination of both types of systems, thus getting specific alarms for those problems which are known to be most common or most costly if left undiscovered, and at the same time get a list of indiscriminant alarms as a way of alerting him to more rare and unforeseen problems. If any practical and/or commercial implementation of some variation of these methods were to be attempted, it would be beneficial to first consult real world farmers to get their input on what type of alarm (specific, indiscriminant, or both) they would prefer to get.

Furthermore, even alarm systems based on the DLM/Cholesky method do not need to operate on an all-or-nothing principle; take for example the problems with metabolic diseases such as ketosis, fatty liver, milk fever, and others, which are common in dairy cows in the first few weeks after calving. These problems are all identifiable by changes in milk yield and appetite, among other symptoms, and the treatments are fairly similar, as they involve giving the cows *e.g.* glucose or calcium orally (LeBlanc 2010; Rajala-Schultz et al. 1999). For this reason, a dairy farmer could apply the DLM/Cholesky method to just the relevant metabolism-related data and thus get semi-specific alarms concerning this collection of metabolic diseases. Demonstrating this application would be an interesting goal for future research.

#### 10.2.3 Naive Bayesian Classifier

In Paper 5, the DLM/NBC method was introduced as a method for detecting a specific problem rather than just any problem, as opposed to what is achieved with the DLM/Cholesky method discussed above.

The ROC curves seen in Paper 5 Figure 5 give a rather convincing impression that the DLM/NBC method is useful for extracting information from the available data, which was valuable for detecting mastitis. Because the focus in Paper 5 was exclusively on detecting mastitis, no other conditions or diseases were considered. In other words, it is not possible, based on the results presented in Paper 5, to say exactly how exclusive the detection of mastitis actually is. Specifically, if there are other diseases which cause some of the same symptoms as mastitis, these events might also raise the mastitis alarm. Ketosis, as an example, is known to result in lowered milk yield (Rajala-Schultz et al. 1999), which is also an important symptom of mastitis (Viguier et al. 2009). This being said, ketosis does not raise the EC or the SCC of the milk the way mastitis does, which illustrates the whole purpose of combining several data sources for specific event detection: if the milk yield is lower than expected, but the EC and SCC are normal, we are probably not dealing with mastitis, and so the mastitis-trained NBC should not raise an alarm, although another, say ketosis trained, NBC might do so. That is the theory, anyway. How event specific the DLM/NBC method actually is will depend on how specific the measurable symptoms of the event of interest are. How mastitis specific the DLM/NBC implementation presented in Paper 5 actually is, and what types of events it will erroneously flag as mastitis, would be interesting questions to answer in a future study.

Just as with the summary/logistic regression method, the NBC part of the DLM/NBC method serves as a classification tool, and could in principle be replaced with any one of several other such tools. Testing various classification tools in combination with various measures of the DLM forecast errors (categorized or numerical) would be a relevant objective for future research.

### 10.2.4 The three methods, summarized

Table 1 provides an overview of which of the three detection/forewarning methods, *i.e.* the three combinations of the pre-processing and classification models, that are useful for handling environment-oriented data, animal-oriented data, and whether or not the method can be used to provide event specific alarms, given the discussions seen above. The label "Yes" signifies that a method has been found to be useful for the relevant purpose, "No" signifies that the method was found not to be useful for the purpose, "(Yes)" signifies that the method should in principle be useful for the purpose, but that this was not the case given how the method was implemented, and "No?" signifies that the method was not tested for the purpose, but that given what else is known, it seems unlikely that it would be useful for it.

Specifically, the summary/logistic regression method has not been tested on animal-oriented data in any of the studies presented in this thesis. It stands to reason, however, that using static summary statistics on animal-oriented data, with no concern for the time series element, would probably not yield very good performances for detection or early warnings about undesired events. This is because animal behavior and physiology will naturally differ between individuals, and by extension between groups of individuals such as a group of pen mates in a pig herd. In contrast, the implicit assumption behind basing alarms on simple summary statistics concerning the animals is that they are all essentially identical. For this reason, much better performances should be expected from learning what would constitute the "normal" state for a specific group or a specific individual, and then monitor the system for deviations from this particular state of normalcy.

Of course there may be special cases where summary values are more useful than time series modeling, as was the case with SCC in the Paper 5. Those daily sensor values were too unreliable to be taken at face value, and were therefore categorized instead, as explained in more detail in the discussion of the DLM/NBC method in section 10.1.2. In a few cases in the scientific literature, summary data have actually been used as the sole input for classification models, such as the raw EC values in milk for detection of mastitis (Nielen et al. 1995b) or the combination of summary statistics (inter-quarter ratios) of EC and raw SCC measurements for detecting mastitis (Mollenhorst et al. 2010). Both of these studies showed less than impressive predictive performances.

**Table 1: Overview of the three main methods used in the papers discussed in this thesis, and whether or not they were found to be useful for three different purposes. Yes = useful. No = not useful. (Yes) = useful in principle, but not in the way it was implemented. No? = it was not tested, but given what else we know, it seems unlikely.**

| Method | Environment oriented data | Animal oriented data | Event specific alarms |
|---|---|---|---|
| Summary/Logistic regression | Yes | No? | (Yes) |
| DLM/Cholesky | No | Yes | No |
| DLM/NBC | Yes | Yes | Yes |

Remember that each of these three alarm methods have only been attempted on data from one herd. Investigating the inter-herd applicability of each of these methods would be a relevant goal of future research. This should be done by training the various models on data from one herd and then test those models on one or more other herds, where the same or comparable data has been collected.

## 10.3 Relative information values

A key point of interest throughout the research presented in this thesis was to determine which of the available data streams were useful for detecting or forewarning about undesired events. In other words, to answer the question: where in the data lies the information? Moreover, do some data streams contain more information than others, and if so, what are the relative information values of the individual data streams? It should be clarified that the term "information value" is only intended to refer to the resulting increase in alarm performance when a data stream is included, and the loss in performance seen when a data stream is omitted. That is to say no monetary considerations have been made when estimating the relative information values in the papers related to this thesis. As pointed out in a review by Cornou and Kristensen (2013), calculating the exact numerical monitory value of livestock information/monitoring systems are difficult for various reasons, and making such estimates is only rarely done in the existing scientific literature. Additionally, the specific monetary concerns were simply beyond the scope of the PhD project presented in this thesis.

In Paper 2, only one type of data was considered, namely pen level temperature data. Nevertheless, because two temperature sensors were placed in each pen and a total of eight derived values were made from these sensors per day, some relevant conclusions can be drawn. First of all, those derived variables which did not prove to be significant ($p < 0.05$) or borderline significant ($p < 0.1$) were excluded from further consideration by backwards elimination, as explained in section 3.2.

Of the five derived variables which remained, as seen in Paper 2 Table 1, three were related to the rate of change in temperature. This suggest that an unstable temperature is an important predictor for the undesired events considered in the paper, which is consistent with existing literature (Lopez et al. 1991b; Lopez et al. 1991a; Shimizu et al. 1978). The two remaining derived variables had to do with the absolute minimum and maximum daily temperatures in the lying area of the pen. Most significantly, higher values of the minimum daily temperature increased the risk of observing the undesired events of interest. This is of course consistent with the knowledge that if the temperature is above the pigs comfort level, the problem with pen fouling increases with increasing temperatures (Aarnink et al. 2006).

Both of the variables relating to the absolute maximum and minimum temperatures came from the sensor in the lying area. This suggest that for the purpose of providing forewarnings about pen fouling specifically, the data collected in the lying area probably has more information value than the data collected by the section corridor.

Conversely, two of the three predictive variables related to the rate of change were collected by the sensor near the section corridor, while the last was collected in the lying area. It should however be remembered that one of the these variables, namely the rate of temperature decrease by the corridor, only showed borderline significance, and that it seemed to have a lowering effect on the probability of observing undesired events, when the rate of temperature decrease became greater. This effect is both counterintuitive and goes against the existing literature as described above. This particular finding should thus be considered with a high degree of skepticism. This also means that no single position, by the lying area or the section corridor, can be said to contain more information than the other when the rate of temperature change is concerned, which would be particularly relevant for providing forewarnings about outbreaks of diarrhea.

All of these relative information interpretations will of course have to be verified rigorously in future research, as part of the previously mentioned effort to maximize the utility of the environment-oriented data.

In Paper 4, the relative information value of the seven included sensor data streams were estimated by a systematic sequence of omission and inclusion scenarios. First, variables were omitted from the model one by one or in meaningful pairs (*i.e.* the two temperature data streams and the two water data streams), and reduction in the performance, compared to including everything, determined the relative information values. Secondly, those variables which had proven their value by omission were sequentially included in the model on their own, or in the same relevant pairs as mentioned above. Here, those single variables which resulted in the best predictive performance were considered to have the highest relative information value. By comparing the predictive performances of all exclusion and inclusion scenarios, the ranked list of information value, which is seen in section 4.4, could be made. This list showed temperature information to contain only very little information, which sparked the discussion seen in section 10.1.1. The two drinking behavior variables were by far the most informative, accounting for almost all of the predictive performance achieved in Paper 4. Madsen and Kristensen (2005) argued that deviations from the pigs' normal diurnal drinking pattern could be used for detecting disease such as diarrhea or the effect of stressors, and the findings in Paper 4 seem to strongly support this idea. In Paper 4 it was further seen that the data streams with low observation frequencies, *i.e.* the weekly live weight measurements and the daily feed amount registrations, proved to have the least information value, suggesting that more frequent observations provide more useful information.

In Paper 5, the information value of the various sensor packages, rather than the data from the individual sensors within those packages, were assessed by systematic omission of one package at a time. The results, which are seen in Paper 5 Table 4, show that both the milk meter and the AfiLab (packages 1 and 2, respectively) provide useful information for detecting mastitis. Given that the milk meter measures the milk yield and the EC, and the AfiLab measures, among other things, SCC, this finding is hardly surprising, as these are probably the three most well documented indicators of mastitis (Viguier et al. 2009). What is more interesting is the synergistic effect seen when combining these two sensor packages, as the performance with this combination is higher than with either of these two packages alone. As SCC is often used as part of the definition for the gold standard of a cow being mastitis positive (which was not the case in the data used in Paper 5), only two studies could be found where somatic cell count were included as a predictive variable (Mollenhorst et al., 2010; Kamphuis et al., 2008) without also being a defining characteristic of mastitis. In both of these studies, SCC was combined with EC, and both studies concluded that this combination improved the detection performance compared to using either EC or SCC alone, and so the synergistic effect observed in Paper 5 is in line with the mentioned literature.

The fact that Paper 5 showed that inclusion of the non-sensor information had a small but consistent positive effect on the performance, as mentioned in section 4.4, is broadly consistent with the findings of Steeneveld et al. (2010). That being said, however, when the sensitivity was held at 0.80, Steeneveld et al. (2010) were able to demonstrate a 40 % reduction in false positive alarms by combining non-sensor data with the information in the alarms raised by the automatic milking system. The effect seen from non-sensor data in Paper 5 is nowhere near that great, as the values in Paper 5 Table 4 will attest to. In fact, including non-sensor data reduced the number of false positives by around 8 % while not affecting the number of true positives (data not shown). It is conceivable that some of the explanation for this difference shall be found in how the non-sensor data were categorized in the two papers. For example, in Paper 5 the progression of the lactation period was categorized into 43 weeks, which was not very informative (data not shown). Steeneveld et al. (2010) on the other hand divided the lactation into periods of 30 days, which they found to be a significant ($p = 0.002$) predictor of whether an alarm was a true or a false positive.

While the relative information value contained in the different data streams may vary, one important lesson was consistently demonstrated in all five papers presented in this thesis: precision data, *i.e.* data (animal-oriented as well as environment-oriented) relating to specific animals at the group or individual level, are important for the purpose of automatic detection of undesired events at those levels.

## 10.4 Performance evaluation

When considering the performances presented in Papers 2 and 4, two things are important to remember.

First of all, only registrations of two types of events were considered in these papers, namely diarrhea and pen fouling. It is nonetheless known that several other problems, including pneumonia and influenza, were present in the herd, but no information on when or where these problems occurred was available when writing Papers 2 and 4. This is probably not a big an issue in Paper 2, even though the logistic regression used in that paper was designed to raise indiscriminant alarms about diarrhea and pen fouling. In spite of this indiscriminant nature of the alarms, the temperature variables which were included as predictive variables were specifically related to the risk of either diarrhea or pen fouling, as discussed several times previously. For Paper 4 on the other hand, the DLM/Cholesky method should raise indiscriminant alarms about any deviations from the normal healthy situation. It is therefore more than likely that at least some of the assigned "false positive" alarms were in fact raised in response to some of the other problems. This means that the specificities reported in Paper 4 are most likely underestimated.

Second, the available registrations of diarrhea and pen fouling were made by the farm staff based on their visual observations of the pens. These registrations were treated as the gold standard for when a true positive event occurred. For pen fouling, these registrations are probably very reliable. For diarrhea, however, Weber et al. (2015) showed that generally for weaned pigs in Denmark, one third of the non-medicated pigs, which were assessed as healthy by the farm staff, did in fact have diarrhea when they were examined clinically. If a similar tendency holds true for finisher pigs, the number of gold standard events will be underestimated. By the same token as above, it would then be very likely that the specificity of the alarms was even further underestimated.

The prediction window method, as described in section 3.6, was chosen as the performance evaluation method in Papers 2, 4, and 5, because it is a commonly used for performance evaluation in PLF studies related to *e.g.* detecting mastitis or estrus in dairy cows (Hogeveen et al. 2010). Remember that this method starts by determining which observation times (days, hours, milkings, etc.) coincide with the observation of an undesired event, and then proceeds to check if any alarms were raised within a certain number of observation times before or after the event was observed. In other words, this method is based on the assumption of exact foreknowledge about when the events will occur, which in reality will never be the case. This means that the forewarning performances reported in most studies, including the ones presented in this thesis, may be overestimated.

The major premises behind this common prediction window approach is that data collected prior to an event observation may contain information which could show that the event was under way, or it may be that the effects of an event cannot be seen in the collected data until sometime after the event has occurred. These are legitimate points, but they need to be addressed in a way which is more consistent with the realistic daily operations on a real world farm. One way of doing this could be to separate what might be called the observation window from the prediction window, as illustrated in Figure 11. The idea behind this alternative evaluation approach is that the observation window is always retrospective while the prediction window is generally prospective and minimally retrospective.

This evaluation approach is closely connected to how the alarms would be raised. Specifically, at any given observation time, the system should look back a number of time steps (*e.g.* hours, days, milkings, etc.). Based on some criteria, *e.g.* the greatest increase and decrease in temperature, or the greatest cumulative sum of DLM forecast errors, the system could either raise an alarm concerning some undesired event or not. This alarm would indicate that the undesired event is expected to occur at some time within the range of the prediction window. If an event then does occur within the prediction window, the alarm is counted as a true alarm. If no event occurs within the prediction window, the alarm would be counted as a false alarm. If the system does not raise an alarm based on the observations in the observation window, and no events are seen in the prediction window, that observation would be counted as a true negative alarm, while if an event did occur within the prediction window, the observation would be counted as a false negative alarm.

The important point is that the observation window and the prediction window are separate entities, even if they do have some overlap, as is the case in the examples seen in Figure 11. This separation means that even if undesired events do occur within the range of the observation window as shown in Figure 11 B, this does not count as a true positive alarm if it is not also covered by the prediction window.

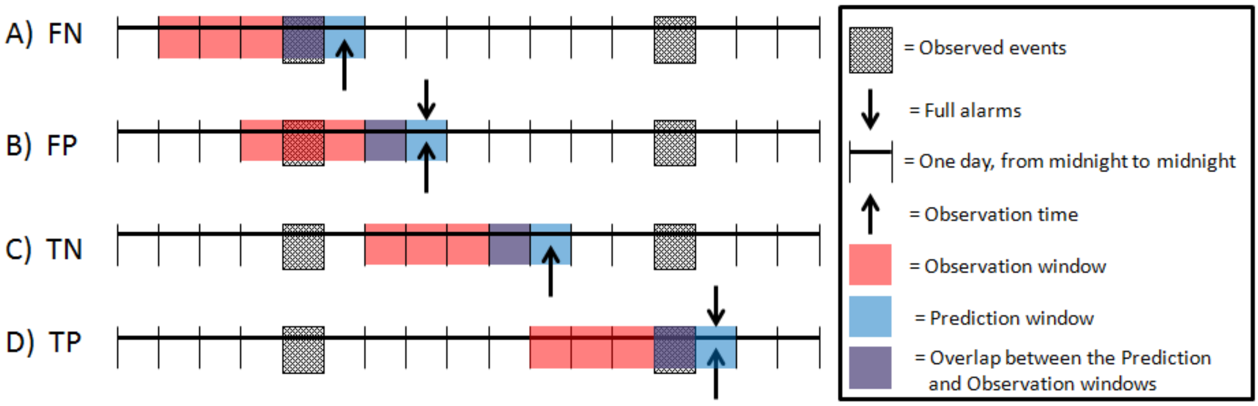



**Figure 11: Illustration of a performance evaluation method, intended to be a more realistic alternative to the prediction window method. The observation window and the prediction window are separate entities, although in this example they have a one day overlap. The observation window is always retrospective, while the prediction window is generally prospective. Both windows move with the observation time. Alarms are raised based on the observations made in the observation window (here four days). A) If no alarm is raised, but an event is observed within the prediction window (here either on the observation day or the day before) it is counted as a false negative (FN). B) If no event is observed within the prediction window, but an alarm is raised, it is counted as a false positive (FP), even if an event is observed within the observation window. C) If no alarms are raised and no events are observed within the prediction window, it is counted as a true negative (TN). If an alarm is raised and an event is observed within the prediction window, it is counted as a true positive (TP).**

This method of raising generally prospective alarms based on retrospective observation windows would make sense in a real world farm scenario. In scenarios such as the ones covered in Papers 2 and 4, where animal and environment oriented data is available on an hourly basis, the system could look back, say 96 hours (as in the example in Figure 11), every hour and thus give an hourly updated alarm list. If desired, the alarm list could also be updated less frequently, say once or twice per day. Getting an updated alarm list twice per day might actually make the most sense from a biological perspective: pigs are known to naturally be most active between 8 AM and 10 AM and again between 3 PM and 5 PM (VSP 2010), which makes these the ideal times for the farm staff to check on the pigs. Thus having the system provide an updated alarm list at either 8 AM, 3 PM, or both, depending on the standard practices and the preferences among the staff, could provide the farm staff with the most updated list of high risk pens before starting the routine herd rounds. The performance of the detections and forewarnings achieved with each of these three alarm list strategies might differ, and therefore it is relevant to have some idea of which of these strategies, or any other option, would be preferred by actual farmers, before optimizing for the performance of any one strategy.

It might also be worth investigating whether the farmers would prefer getting retrospective or prospective alarms. Retrospective alarms would mean that the farmer would generally get the alarms when an event such as diarrhea or pen fouling had already occurred. The upside to this is that the problem would be visible, and so the farmer would be within his right to administer various treatments such as antibiotics. The drawback is that in this scenario, the problem might already have been noticed by the farm staff, in which case the alarm would just bring old news. The prospective alarms would by definition mean that the animals would not yet show symptoms of *e.g.* diarrhea when the farmer received the alarms, and the farmer would thus not be allowed to administer antibiotics anyway. That being said, unpublished data[2] show that treating weaned pigs after intestinal infection, but before symptoms are visible, led to an overall increase in daily weight gain among the pigs. Therefore, if a non-invasive forewarning system could be shown to be sufficiently reliable, early warnings of diarrhea might be a valid reason for the farmer to have relevant fecal samples analyzed in order to determine whether or not a pen of pigs have a still un-symptomatic intestinal infection, which would then legitimize *e.g.* antibiotic treatment.

If one of these implementations of the alternative alarm strategy were to be decided upon, it would be relevant to study how the alternative strategy compares to an implementation of the common prediction window strategy, in order to determine if detection and forewarning performances are currently being systematically overestimated in the literature. To do this, the two implementations of the two different methods would have to be as comparable as possible. For example, an implementation of the alternative evaluation method, where the observation window and the prediction window both cover only the day being observed, would be equivalent of a -0/+0 prediction window of the common prediction window method, and they should yield the same performance.

If the observational time unit were one day and the -3/+1 prediction window were to be used with the common method, as was done in *e.g.* Paper 2, this would mean that any alarms raised up to three days before or up to one day after an undesired event would count as a true positive alarm. The best approximation using the alternative method described here, would be one with an observation window stretching from four days to one day before the current observation time, and a prediction window stretching from one day before the current observation time until the current observation time. This is the same scenario as has been illustrated in Figure 11. In this scenario, an alarm can be raised up to one day after an undesired event has occurred and still count as a true positive, as is the case in the -3/+1 implementation of the common method, and the data collected up to four days before the observation time can be included in deciding whether or not to raise an alarm.

## 10.5 Summary of research perspectives

The discussions presented in this chapter raised a number of research questions which should be addressed in future research. For the benefit of the reader, these research perspectives have been sorted and summarized here.

### 10.5.1 Farmer preferences

To increase the chance that an alarm system will actually be useful in practice, real-world farmers should be consulted to determine the nature of the alarm system they would prefer to use. The relevant questions to ask the farmers are:

[2] Personal communication, Nicolai Weber, veterinarian and PhD student at the University of Copenhagen, Department of Large Animal Sciences

What is the preferred frequency for updating the list of alarms:

a. One list of alarms per day?
b. Two lists of alarms per day?
c. Hourly updated alarm lists?
d. Some other frequency?

What type of alarms would be preferable:

a. Indiscriminant alarms (as with the DLM/Cholesky method)?
b. Event specific alarms (as with the DLM/NBC method)?
c. A combination of both?

What temporal orientation should the alarms have:

a. Retrospective? (when the farmer gets the alarm, the event has already occurred)
b. Prospective? (if the farmer gets an alarm, he can expect that an events will be observed within a specified amount of time in the future)

#### 10.5.2 Realistic performance evaluations

A more realistic performance evaluation should be implemented. This method should correspond to the kind of alarms which would be useful in a practical herd setting and. This evaluation method should be tested against the prediction window method, which is currently commonly used.

#### 10.5.3 Method improvements

The methods for extracting information related to event specific detections and forewarnings could conceivably be improved by applying alternative data handling and classification tools to the same overall method frameworks which have been presented in the papers discussed in this thesis.

For environment-orientated data, future studies should focus on:

a. Including summary statistics of humidity in addition to temperature.
b. Account for the fact that optimal comfort temperature will vary with the pigs' growth.
c. Evaluate the performance when using alternatives to logistic regression for combining the summary statistics, such as artificial neural networks, support vector machines, decision trees, and others.
d. Evaluate the performance achieved when the environment-oriented data are monitored by a DLM, which is optimized specifically and solely for the environment-oriented data.

For animal-oriented data, future studies should focus on:

a. Testing the use of different measures of the DLM forecast error, such as various error classification cutoffs, numeric values, and cumulative sums of the forecast errors.
b. Testing different classification tools for unifying the forecast errors, such as artificial neural networks, support vector machines, decision trees, and others.

### 10.5.4 Method verifications

The methods for detections and early warnings described in the papers presented in this thesis, as well as any improved version derived from those methods, would need to undergo further verifications before any commercial implementations can be considered. These verifications should:

a. Compare the best performances achieved when monitoring the environment-oriented data with a DLM strategy compared to with a summary strategy. This is to test the finding that environment-oriented and animal-oriented data are best handled in different ways.
b. Test the between-herd applicability of the event specific and the indiscriminant detection/forewarning methods.
c. Determine the degree to which the event specific methods are able to distinguish between separate events.
d. Demonstrate the utility of the DLM/Cholesky method for semi-specific detection of metabolic diseases in dairy cattle.

## 10.6 Chapter 10 references


Andersen, H.M.L. et al., 2008. The ear skin temperature as an indicator of the thermal comfort of pigs. *Applied Animal Behaviour Science*, 113(1-3), pp.43–56.

Cavero, D. et al., 2008. Mastitis detection in dairy cows by application of neural networks. *Livestock Science*, 114(2-3), pp.280–286.

Cornou, C. & Kristensen, A.R., 2013. Use of information from monitoring and decision support systems in pig production: Collection, applications and expected benefits. *Livestock Science*, 157(2-3), pp.552–567.

Cornou, C. & Lundbye-Christensen, S., 2012. Modeling of sows diurnal activity pattern and detection of parturition using acceleration measurements. *Computers and Electronics in Agriculture*, 80, pp.97–104.

Hogeveen, H. et al., 2010. Sensors and clinical mastitis--the quest for the perfect alert. *Sensors (Basel, Switzerland)*, 10(9), pp.7991–8009.

Kamphuis, C. et al., 2008. Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. *Journal of dairy science*, 91(12), pp.4560–4570.

Kamphuis, C. et al., 2010. Decision-tree induction to detect clinical mastitis with automatic milking. *Computers and Electronics in Agriculture*, 70, pp.60–68.

LeBlanc, S., 2010. Monitoring metabolic health of dairy cattle in the transition period. *The Journal of reproduction and development*, 56 Suppl, pp.S29–S35.

Lopez, J. et al., 1991a. Effects of temperature on the performance of finishing swine : II . Effects of a cold , diurnal temperature on average daily gain , feed intake , and feed efficiency. *Journal of animal science*, 69, pp.1850–1855.

Lopez, J. et al., 1991b. Effects of temperature on the performance of finishing swine: I. Effects of a hot, diurnal temperature on average daily gain, feed intake, and feed efficiency. *Journal of animal science*, 69(5), pp.1843–1849.

Madsen, T.N., Andersen, S. & Kristensen, A.R., 2005. Modelling the drinking patterns of young pigs using a state space model. *Computers and Electronics in Agriculture*, 48(1), pp.39–62.

Madsen, T.N. & Kristensen, A.R., 2005. A model for monitoring the condition of young pigs by their drinking behaviour. *Computers and Electronics in Agriculture*, 48(2), pp.138–154.

De Mol, R.M. et al., 2001. Detection of estrus and mastitis: Field performance of a model. *Applied Engineering in Agriculture*, 17(3), pp.399–407.

De Mol, R.M. et al., 1997. Results of a multivariate approach to automated oestrus and mastitis detection. *Livestock Production Science*, 48, pp.219–227.

Mollenhorst, H., van der Tol, P.P.J. & Hogeveen, H., 2010. Somatic cell count assessment at the quarter or cow milking level. *Journal of dairy science*, 93(7), pp.3358–3364.

Nielen, M., Spigt, M., et al., 1995. Application of a neural network to analyse on-line milking parlour data for the detection of clinical mastitis in dairy cows. *Preventive Veterinary Medicine*, 22(94), pp.15–28.

Nielen, M., Schukken, Y.H., et al., 1995. Comparison of analysis techniques for on-line detection of clinical mastitis. *Journal of dairy science*, 78, pp.1050–1061.

Ostersen, T., Cornou, C. & Kristensen, A.R., 2010. Detecting oestrus by monitoring sows' visits to a boar. *Computers and Electronics in Agriculture*, 74(1), pp.51–58. Available at:

Rajala-Schultz, P.J., Gröhn, Y.T. & McCulloch, C.E., 1999. Effects of milk fever, ketosis, and lameness on milk yield in dairy cows. *Journal of dairy science*, 82(2), pp.288–294.

Shimizu, M., Shimizu, Y. & Kodama, Y., 1978. Effects of ambient temperatures on induction of transmissible gastroenteritis in feeder pigs. *Infection and Immunity*, 21(3), pp.747–752.

Steeneveld, W. et al., 2010. Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems. *Journal of dairy science*, 93(6), pp.2559–2568.

Viguier, C. et al., 2009. Mastitis detection: current trends and future perspectives. *Trends in biotechnology*, 27(8), pp.486–93.

VSP, 2010. Manual om Vækstmanagement. Available at: http://vsp.lf.dk/~/media/Files/PDF - Viden/Til staldgangen/Haandbogsblade/Vaekstmanagement/H2 - Daily supervision of finishers.pdf [Accessed February 14, 2016].

Weber, N. et al., 2015. Occurrence of diarrhoea and intestinal pathogens in non-medicated nursery pigs. *Acta Veterinaria Scandinavica*, 57(1), p.64.

West, M. & Harrison, J., 1997. *Bayesian Forecasting and Dynamic Models* 2nd ed., New York, USA: Springer.

Aarnink, A.J.A. et al., 2006. Temperature and body weight affect fouling of pig pens. *Journal of Animal Science*, 84, pp.2224–2231.

# Chapter 11: Conclusions

The results of the studies presented in this thesis clearly demonstrate that precision data, meaning frequently collected data pertaining to specific animals within the herd, are important for the purpose of identifying specific groups or individuals with problems relating to health and welfare.

The results further show that multivariate DLMs are effective tools for combining the very diverse data streams which will have to be collected for the purpose of model-based monitoring of animal production herds. Specifically the DLM can be used to meaningfully co-model data which differ considerably in source, numerical magnitude, and observational frequency. In principle, this includes environment-oriented data as well as animal-oriented data, although the results presented in this thesis suggest that these two data types would need to be monitored separately in order to achieve the best performances of detections and forewarnings.

It was further shown that the multivariate DLM offers a simple way to establish the relative information values of multiple data streams. This is generally achieved by designing a multivariate DLM to co-model all variables of interest, and then systematically omit and include selected variables from the model while measuring the resulting performance of the alarms produced by said model.

As demonstrated in several papers in this thesis, a DLM-based alarm system will raise the alarms based on the forecast errors produced by the DLM at each observation step. There are several different methods by which this could be accomplished. The papers described in this thesis presented two such methods. The first method was forecast error unification via Cholesky decomposition. In this case, the alarms would be raised if the value of the unified error surpassed a set threshold. The second method was a categorization of each of the forecast errors at each observation step, based on the direction and magnitude of the error. In that case, a probability of observing an undesired event was calculated with a naïve Bayesian network, and an alarm was raised if this probability surpassed a set threshold.

An additional important conclusion is that evaluating these performances is in no sense a trivial matter, and that the currently common prediction window method is not optimal, as it is too far removed from the practical reality of monitoring a herd.

Although the results presented in this thesis are generally encouraging, a commercial implementation of the described methods would be premature at the present time. Rather, the results should be seen as a source of inspiration for a varied collection of future research. The overall goals of this future research should be to build and improve upon the methods described in this thesis, as well as to verify their utility in multiple different herds, representing various realistic farming practices.