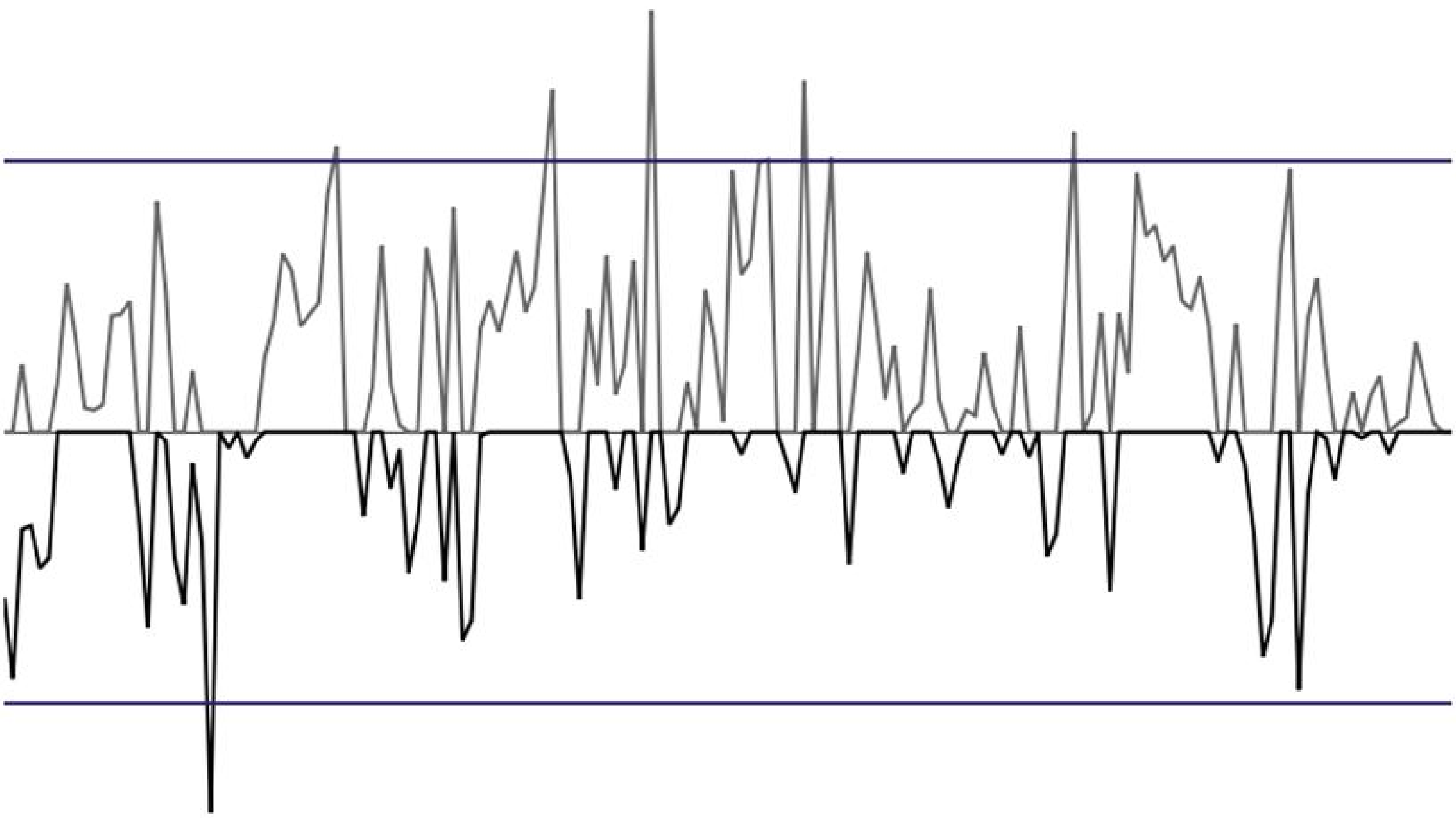
FACULTY OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

Spatial modeling of drinking patterns as a tool for reducing alarms in pig production

KATARINA NIELSEN DOMINIAK

PhD thesis · 2017

This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen



Spatial modeling of drinking patterns as a tool for reducing alarms in pig production

PhD Thesis

Katarina Nielsen Dominiak

HERD

Centre for Herd-oriented Education, Research and Development

Department of Veterinary and Animal Sciences

University of Copenhagen

August 2017

Principal supervisor
Professor Anders Ringgaard Kristensen
Department of Veterinary and Animal Sciences
University of Copenhagen, Denmark

Co-supervisor
Senior Researcher Lene Juul Pedersen
Department of Animal Sciences - Behaviour and stressbiology
Aarhus University, Denmark

Assessment committee
Associate Professor Matt Denwood (Chair)
Section for Animal Welfare and Disease Control
University of Copenhagen, Denmark

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

Associate Professor (Senior lecturer) Tomas Norton
Division Animal and Human Health Engineering
Katholieke Universiteit Leuven, Belgium

This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen 27/08/2017

Spatial modeling of drinking patterns as a tool for reducing alarms in pig production

“Jump, and you will find out how to unfold your wings as you fall”

- Ray Bradbury

To Henning, Gustav and Ebbe,
I love you to the Moon and back

ACKNOWLEDGEMENTS

First and foremost I would like to express a heartfelt thank you to my primary supervisor, Anders Ringgaard Kristensen for having the faith in me to begin and complete such a project despite many years out of academia. Thank you also to my co-supervisor, Lene Juul Pedersen, for taking this role and for inviting me to have an everyday workplace amongst wonderful colleagues in the ASB group at Foulum.

I would also like to give a huge thank you to Mikkel Boel, who encouraged me on a daily basis during the writing of my thesis. Likewise a huge thank you to Jeff Hindsborg, whom I had the pleasure of working with on two of my manuscripts, and to Dan Børge Jensen, and Anna Helena Stygar, who are wonderful colleagues and traveling partners. A warm thank you also to all colleagues at Grønnegaardsvej. You have always made me feel welcome whenever I was in Copenhagen.

Furthermore, I would like to acknowledge all the wonderful people involved in the PigIT project, both at the University of Copenhagen, at the Pig Research Centre, and at Aarhus University, Foulum. It has been a pleasure to work with you as a PhD student and as project coordinator. Thank you.

It should, of course, be noted, that conducting this PhD only was possible due to the funding by the Danish Council for Strategic Research (The PigIT project, Grant number 11-116191).

A heartfelt thank you to Kaj, Anne Marie and Thomas for trusting a newly hatched biologist to manage a sow herd with weaners eleven years ago. It was the perfect way to be introduced to the always fascinating world of pigs.

Last, but not least, I would like to thank family and friends for their love, support, and understanding throughout my PhD study. Thank you Mom and Dad, and in particular, thank you my wonderful husband, Henning, for loving me and bearing with me through the very intense periods of my study together with our two fantastic sons, Gustav and Ebbe. I love you all more than words can say.

ABSTRACT

The aim of this PhD thesis was to investigate whether simultaneous sensor-based monitoring of drinking patterns in multiple pens across a herd of growing pigs, could be used to detect outbreaks of diseases in specific areas of a herd of growing pigs. The thesis is a contribution to the collection of scientific work of the international PigIT alliance, which has the overall research goal to improve animal welfare and productivity in Danish production of growing pigs, using advanced ICT methods.

Despite a generally high health and welfare status in Danish pig production, outbreaks of diarrhea and fouling, which is a change in behaviour where the pigs start to lie on the slatted area of the pen and excrete in the lying area, constitute an everyday challenge in the herds.

The structure of a Danish herd for growing pigs is sectionalized in the way that a herd consists of a number of identical sections, and each section consists of a number of identical pens. Such a sectionalized structure of a herd enables the use of a spatial model, which can relate specific observations to specific areas in the herd.

Thus, in this study, water sensors were placed in multiple pens within multiple sections in two herds of growing pigs (weaners 7-30 kg, finishers 30-110 kg), and the specific hypotheses defined in this PhD study were: *Hypothesis I)* drinking patterns between pens within a section and sections within a herd of growing pigs are correlated, and this correlation can be modeled using model parameters defined at different spatial levels, and *Hypothesis II)* changes in the drinking patterns of growing pigs are influenced by diarrhea and fouling. By monitoring the water consumption simultaneously in multiple pens and sections, outbreaks of the conditions can be detected in specific areas.

In Paper I, an extensive literature review, was conducted. The objective of the review was to provide an overview of different methods for reducing or prioritizing alarms from sensor-based detection models in livestock production. The performances of detection models, developed over a twenty-years period, were furthermore compared to three criteria in order to assess their implemental value in a commercial herd. The results of Paper I showed that only three methods were developed for reducing or prioritizing false alarms. In addition, the results showed that none of the evaluated models were suited for implementation in a commercial herd. Poor detection performance was the primary cause for the models being unsuited for implementation.

Based on the literature review, further research is needed on new approaches for improving performances and reducing alarms from sensor-based detection models in livestock production.

In Paper II, Hypothesis I was addressed, and a spatial model was developed. The simultaneously monitored drinking patterns were modeled by a multivariate *dynamic linear model*, where each monitored drinking pattern constituted a unique variable. Hereby all monitored drinking patterns were modeled simultaneously, and different correlation structures could be defined. Thus, seven different model versions were defined to express different degrees of correlation between the drinking patterns.

Subsequently their ability to fit the data was measured as *mean square error* (MSE). The results indicated a correlation in data from pens within the same section for the finisher herd (MSE = 13.850). For the weaner herd, the results indicated an inverse relation between the model fit and the degree of correlation, and the best fitting model version (MSE = 1.446) therefore expressed the lowest degree of correlation between drinking patterns.

Based on the findings in Paper II, there is a degree of correlation between the drinking patterns in different areas of a herd. However, the results for the weaner herd indicated that there were too few pigs in each pen to evaluate the model rightfully, and an external validation of the model would be a first step in identifying how future work on the model should be conducted.

In Paper III, Hypothesis II was addressed. The seven model versions, developed in Paper II, were evaluated for their abilities to detect outbreaks of diarrhea or fouling in either a specific pen, a specific section, or any pen in the herd. The evaluation was conducted by applying a *two-sided tabular Cusum control chart* to generate alarms from the output of the models. The accuracy of the alarms were then evaluated given three lengths of detection windows. The results were reported as the *the area under the curve* (AUC), and for both herds, the longest detection window combined with the strongest degree of correlation detected events in a specific setion with the highest performances (AUC = 0.98, AUC = 0.94). However, the settings applied to generate these high performances, showed to be of little managerial value. It was found that the same model version combined with the medium-length detection window, was able to detect event in a specific section as well, and contituted better suited setting for implementation. Different postprocessing methods for reducing or prioritizing alarms generated by the Cusum, were furthermore suggested in Paper III, and the potential of an alarm-reducing method was presented by an exploratory example in Paper IV.

Based on the findings in Paper III, it is possible to detect outbreaks of either diarrhea or fouling, and to generate area-specific alarms. However, too many false alarms were generated, and it is suggested that future focus on improving the detection system is targeted at **a)** improving model performances, **b)** methods for prioritizing or reducing the alarms, and **c)** methods for distinguishing between different causes of alarms.

In conclusion, the research presented in this PhD thesis, emphasizes the general challenges in obtaining high detection performances for the detection of specific events in livestock production. Especially the use of indirect indicators for the events of interest impedes high performances. The presented research points out difficulties in using detection performance as an indication for the implemental value of a model in a commercial herd, and suggests that the results of external validation should be used as an indication instead.

SAMMENDRAG

Formålet med denne PhD afhandling var at undersøge, om det er muligt at påvise sygdomsudbrud i specifikke områder af en svinebesætning[1] på baggrund af overvågning af grisenes drikkemønstre i flere stier. Afhandlingen bidrager til det samlede videnskabelige arbejde udført i det internationale forskningssamarbejde, PigIT, der har som overordnede målsætning at øge dyrevelfærden og produktiviteten i danske svinebesætninger ved hjælp af avancerede teknologiske løsninger.

Til trods for den generelt høje sundhed og velfærd i danske svinebesætninger, udgør udbrud af diarré og stivending (en adfærdsændring, hvor grisene "vender" stien, og begynder at gøde i lejearealet og lægge sig i gødearealet) en tilbagevendende udfordring i den daglige drift.

Svinebesætninger i Danmark består typisk af én bygning, der er opdelt i flere identiske sektioner. Derudover består hver sektion af flere identiske stier. Denne sektionsopdelte struktur gør det muligt at anvende en spatiel, eller rumlig, model, som kan forbinde en specifik observation til et specifikt område af besætningen.

Derfor er der, i denne PhD, placeret vandsensorer i flere stier fordelt over flere sektioner i to besætninger med grise i vækst (smågrise, 7-30 kg, slagtesvin 30-110 kg), og to specifikke hypoteser er formuleret som: *Hypotese I)* drikkemønstre mellem stier indenfor samme sektion, og mellem sektioner i samme besætning, er korrelerede og denne korrelation kan udtrykkes i en model, hvis parametre er defineret på forskellige rumlige niveauer, og *Hypotese II)* ændringer i vækstgrises drikkemønstre påvirkes af diarré og stivending. Gennem samtidig sensorovervågning af drikkemønstre i flere stier med grise, kan udbrud af disse uønskede tilstande forudsiges i specifikke områder af besætningen.

Artikel I præsenterer en omfattende litteraturgennemgang. Formålet med litteraturgennemgangen var at skabe et overblik over de metoder, der hidtil er anvendt til at reducere, eller rangordne, alarmer fra sensorbaserede alarmsystemer i husdyrproduktionen. Derudover blev performance[2] af sensorbaserede alarmsystemer, udviklet gennem en tyveårig periode, holdt op mod tre kriterier for at vurdere deres værdi for en kommerciel svineproduktion. Resultatet af Artikel I viste, at der kun var beskrevet tre metoder til at reducere eller rangordne alarmer. Derudover viste resultatet, at ingen af de alarmsystemer, der var beskrevet i den gennemgåede litteratur, egnede sig til brug i kommercielle besætninger. Den primære årsag til den manglende egnethed var dårlig performance i forhold til at påvise udbrud af uønskede tilstande hos dyrene korrekt.

På baggrund af denne litteraturgennemgang kan det fastslås, at yderligere forskning er nødvendig for at øge performance og udvikle nye tilgange og metoder til at reducere mængden af alarmer fra sensorbaserede alarmsystemer i husdyrproduktionen.

Artikel II adresserer Hypotese I og beskriver udviklingen af en rumlig model. De simultant overvågede drikkemønstre fra stierne blev modeleret i en multivariabel *dynamisk linær model*, hvor hver

[1] Besætning med vækstgrise: Smågrise (7-30 kg), eller slagtesvin (30-110 kg)

[2] Præstation, ydeevne, resultat

enkelt drikkemønster indgik som en unik variabel. Herved blev samtlige drikkemønstre modeleret simultant, og forskellige grader af korrelation mellem dem kunne defineres. Syv forskellige modelversioner blev derfor defineret for at udtrykke forskellige grader af korrelation mellem drikkemønstrene. Efterfølgende blev hver modelversion vurderet i forhold til, hvor godt den passede til de observerede data, og graden af tilpasning til data blev udtrykt som *mean square error* (MSE). Resultaterne indikerer, at data er korreleret mellem stier indenfor samme sektion i slagtesvinebesætningen (MSE = 13.850). I smågrisebesætningen indkerer resultaterne derimod, at der er en omvendt relation mellem modellens evne til at passe til data, og den grad af korrelation, der er udtrykt i modellen. Det betyder, at den model, der tilpasser sig data bedst (MSE = 1.446) indikerer, at der er den lavest mulige grad af korrelation mellem de enkelte drikkemønstre.

Resultaterne i Artikel II viser, at der forefindes en grad af korrelation mellem drikkemønstre i forskellige områder af en besætning. De resultater, der blev fundet for smågrisebesætningen, tyder dog på, at der var for få grise i de enkelte stier til at danne grundlag for at vurdere modellens performance på dette datasæt. En validering af modellen på et datasæt fra en anden besætning vil derfor være første skridt i at undersøge, hvordan fremtidig videreudvikling af modellen skal foregå.

Artikel III adresserer Hypotese II, og evaluerer de syv modelversioner, defineret i Artikel II, i forhold til deres evne til at påvise udbrud af diarré eller stivending i en specifik sti, en specifik sektion eller en hvilken som helst sti i besætningen. Evalueringen foregår ved at anvende en *two-sided tabular Cusum control chart* til at danne alarmer udfra data fra hver af de syv modelversioner. Alarmernes nøjagtighed blev vurderet indenfor tre definerede tidsperioder mellem alarm og udbrud af en hændelse. Resultaterne blev angivet som *area under the curve* (AUC), og de viste at for begge besætninger var det kombinationen af den længste tidsperiode, stærkeste grad af korrelation samt påvisning af en hændelse i en specifik sektion, der gav de højeste performances (AUC = 0.98, AUC = 0.94). Denne kombination viste sig dog at ville have meget lille værdi i en kommerciel besætning. Derimod ville samme model, kombineret med den mellemlange tidsperiode, have en langt højere værdi i en kommerciel besætning. I Artikel III blev forskellige metoder til at reducere eller rangordne de alarmer, der blev dannet af Cusum foreslået. Potentialet for en af de alarmreducerende metoder blev derudover illustreret i et eksplorativt eksempel i Artikel IV.

Resultaterne i Artikel III viser, at det muligt at påvise udbrud af enten diarré eller stivending i et specifikt område i en besætning. Der bliver dog dannet for mange falske alarmer, og det foreslås derfor, at fremtidig forbedring af alarmsystemet fokuserer på **a)** at forbedre performance, **b)** metoder til at rangordne, eller reducere mængden af alarmer og **c)** metoder til at skelne mellem forskellige årsager til alarmerne.

Den overornede konklusion af den forsking, der præsenteres i denne PhD afhandling, fremhæver de generelle udfordringer, der ligger til hinder for at opnå høj performance i påvisning af specifikke hændelser i husdyrproduktion. Det er i høj grad anvendelsen, af indirekte indikatorer for de enkelte hændelser, der hindrer høj performance. I det præsenterede arbejde udpeges nogle udfordinger ved at bruge performance som indikation for, hvorvidt et alarmsystem kan give værdi i en besætning. I stedet foreslås det, at en sådanne nytteværdi af et alarmsystem skal vurderes ved at evaluere på datasæt fra andre besætninger.

TABLE OF CONTENTS

OUTLINE OF THE THESIS

Chapter 1 contains an introduction to the present thesis and provides both the background and the state-of-the-art for the conducted research

Chapter 2 presents the scientific aim of the thesis.

Chapter 3 presents the materials and methods used for the three manuscripts in Chapters 6, 7, and 8.

Chapter 4 presents findings and discussions of the four manuscripts.

Chapters 5, 6, 7, and 8 each contain a manuscript conducted during the PhD project

- K. N. Dominiak and A. R. Kristensen. Prioritizing alarms from sensor-based detection models in livestock production - A review on model performance and alarm reducing methods. *Computers and Electronics in Agriculture; 2017.133:46-67*, 2017.
- K. N. Dominiak, L. J. Pedersen, and A. R. Kristensen. Spatial modeling of pigs' drinking patterns as an alarm reducing method. I. Developing a multivariate dynamic linear model. *Submitted for journal publication*, 2017.
- K. N. Dominiak, J. Hindsborg, L. J. Pedersen, and A. R. Kristensen. Spatial modeling of pigs' drinking patterns as an alarm reducing method. II. Application of a multivariate dynamic linear model. *Submitted for journal publication*, 2017.
- K. N. Dominiak, J. Hindsborg, L. J. Pedersen, and A. R. Kristensen. Reducing alarms and prioritising interventions in pig production by simultaneous monitoring of water consumption in multiple pens. Accepted for publication in *Proceedings for the EC-PLF conference in Nantes, 12-14 September 2017*

Chapter 9 contains a general discussion and perspectives of the research presented in the thesis

Chapter 10 provides the overall conclusions of the presented PhD

1

INTRODUCTION

1.1 BACKGROUND

Pig production has changed over the past 40 years. From small scale farms housing both breeders, weaners and finishers, into larger centralized herds, which are highly specialized production units each designed for either breeders, weaners or finisher pigs (Kashiha et al., 2013; Sorensen et al., 2010).

This multi-site production is a part of a bio-security strategy, which has effectively reduced the impact and spreading of diseases and improved the overall health of pigs. Particularly the focus on high health status herds with Specific Pathogen Free (SPF) pigs and All-In-All-Out (AIAO) management has improved animal health in Denmark as compared to other parts of the world (Cameron, 2000; Danish Agriculture and Food Council, 2010).

Despite the high health status of pigs in Danish production herds, some diseases still have significant impact on the welfare and result in too high mortality and reduced productivity. The increased number of animals within the herds leaves less time available to attend to individual pigs during the daily check. It is therefore more difficult for the caretaker to recognize events like diarrhea, tail biting, and fouling, which is a change in behaviour where the pigs start to lie on the slatted area of the pen and excrete in the lying area, at the early signs of an outbreak. Interventions implemented after an outbreak will often reduce the consequences of the condition, but both welfare and productivity will be compromised to some extent for the rest of the growing period of the affected animals.

If the pigs were monitored around the clock, any early signs of undesired events might be recognized. Hereby interventions could be implemented timely enough to either prevent the events from occurring, or limit the consequences to a greater extent than with subsequent interventions.

Constant monitoring by personnel is, however, not a realistic option considering both monetary expenditures and efficiency. Technical solutions are still getting more affordable, though, and by installing sensors in the herd, the pigs would be monitored automatically around the clock, and constant data streams could be modeled into early warning systems.

An early warning system can detect early stages of events like changes in behaviour or outbreaks of diseases. If an event is detected, the warning system will communicate an alarm to the caretaker. The alarms provide the caretaker with valuable information on the status of the animals and will act as decision support in the daily management.

The research, which will be presented here, focuses on the development of such an early warning system. The described PhD project is a part of the larger PigIT alliance[1], which is an international cooperation of scientists working on developing monitoring and decision support tools in growing pigs. By integrating Information and Communications Technology (ICT), PigIT aims to improve both welfare and productivity in Danish production of growing pigs.

PigIT focuses on three events of interest, which have severe impact on both welfare and productivity in Danish pig herds; diarrhea, tail biting and fouling. Tail bites were, however, not occurring in the data included in the present work, and will not be discussed any further in this thesis.

1.2 STATE-OF-THE-ART RESEARCH

The development of sensor-based detection models for livestock production has been a field of increased scientific focus for more than twenty years (Berckmans, 2014; Wathes et al., 2008).

The overall concept of sensor-based detection models is to automatically detect a given condition based on continuous real-time monitoring by one or more sensors. The data observed by the sensors serve as input to one or more mathematical models in a detection system, and if a predefined threshold is exceeded by the modeled data, an alarm is generated. Based on the alarms, the farmer can decide to target interventions and managerial focus, thus using the detection model for decision support in the everyday production.

Throughout the years, livestock science detection models in general, have focused primarily on detecting individual animals with specific conditions like clinical mastitis, lameness and oestrus. Few detection models have been developed for groups of animals, Madsen et al. (e.g. 2005) developed a model for prediction of diarrhea in a section of weaner pigs, and recently Jensen et al. (2017) and Jensen et al. (2016) focused on events in a pen of finisher pigs.

If a warning system could provide precise information on specific animals or groups of animals in need of high attention, it would be of high managerial value to the farmer. Such precise alarms would require high performance of the detection model in order to avoid false alarms. False alarms are both costly and time consuming for the farmer, and in addition to the costs, excessive alarms devaluate the managerial value, and diminishes the trust in the warning system. In other words, a warning system, which generates too many false alarms, will not be suited for decision support and should not be implemented in a livestock producing herd (Hogeveen et al., 2010).

Alarms are generated by comparing the modeled data to a predefined threshold, which is defined according to the nature of the detection system and the condition sought detected. A simple system might measure the body temperature of an animal and generate an alarm if a certain temperature is reached or exceeded, whereas a more complex system might predict the next value of the observed variable and generate an alarm if the modeled data deviates systematically from what is expected.

Whether an alarm is true or false is determined by comparing the time of each alarm with information on whether the condition, or event of interest, occur at the same time or not. For this purpose,

[1] http://pigit.net

the *gold standard* must be known. The gold standard ideally expresses the true state of the modeled system, hence stating with certainty whether the event is present or absent. In practice, however, the gold standard often consists of human observations, which will always be associated with a natural subjectivity, as discussed in Paper II. Although the subjectivity usually is sought minimized by a clear definition of case vs non-case, there is little or no consensus in such definitions across studies of the same condition. Thus in the scientific literature, as different clinical mastitis definitions as "Somatic Cell Count (SCC) above 100,000 cells/ml or treatment performed" (Cavero et al., 2007) and "one or more alerts given in a defined period around the recorded date of an observed case" (Mol and Ouweltjes, 2001) can be found.

Since alarms seldom occur at the exact same moment as the events are registered, periods of time, known as *time windows*, relative to an event, are often defined (Hogeveen et al., 2010; Kamphuis et al., 2010a; Sherlock et al., 2008). All alarms generated within a time window are treated as one single alarm correctly identifying the event. The length of the time window may vary according to the event of interest, and it can extend from before an event is observed to after an event is observed (Cavero et al., 2006; Jensen et al., 2017; Mol et al., 1997).

Whether an alarm is associated with an event or not, lays to ground for a categorization into one of four categories of true and false alarms as follows:

- True Positive (TP) is an alarm occurring in the defined time window around an event
- False Positive (FP) is an alarm occurring outside a time window around an event
- True Negative (TN) is when there is no alarm and no event occurring
- False Negative (FN) is when there is no alarm at an event or during the time window.

As illustrated in Figure 1.1 from Paper III, the length of a time window influences the categorization of true and false alarms, and this will affect the performance of the detection model.

Based on the categorization of the alarms, the model performance is measured by the conditional probabilities *sensitivity* (Se) and *specificity* (Sp), which are estimated as:

$$\mathsf{Se} = \frac{\mathsf{TP}}{(\mathsf{TP} + \mathsf{FN})} \tag{1}$$

and

$$\mathsf{Sp} = \frac{\mathsf{TN}}{(\mathsf{TN} + \mathsf{FP})} \tag{2}$$

where TP denotes the total number of TP cases and accordingly for the other variables, as shown in Paper III.

Sensitivity reflects the model's ability to correctly identify the occurrence of the event of interest, whereas specificity reflects the model's ability to correctly identify the absence of the event of interest. Thus a high sensitivity is needed in order to identify the events, but since the event of interest usually is of low prevalence in the herd, a high specificity is crucial in order to reduce the number of false alarms.

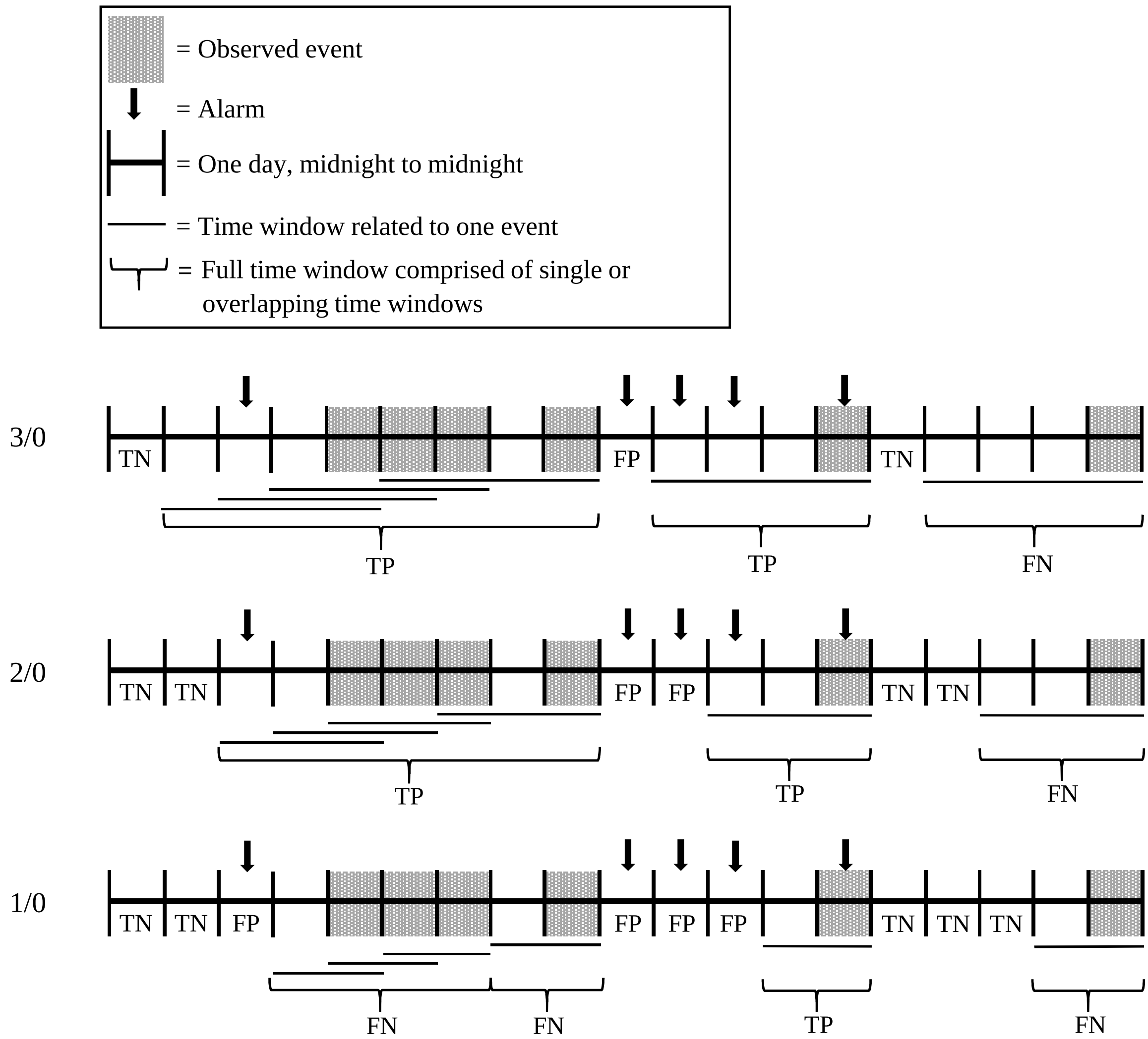


Figure 1.1: Example of definitions of true positives (TP), false positives (FP), true negatives (TN), and false negative (FN). All observed events are associated with a time window, and overlapping time windows are merged into longer windows. Three lengths of time windows are illustrated; 3/0 = three days before an event and zero days after, 2/0 = two days before an event and zero days after, 1/0 = one day before an event and zero days after. All alarms occurring within a time window are counted as one TP alarm. If no alarms occur within a time window, it is counted as one FP. Days outside of time windows but with alarms, are counted as FP, whereas days outside of time windows with no alarms are counted as TN. Based on illustration by Jensen et al. (2017) (Figure from Paper III).

The importance of correct identification of both cases and non-cases has resulted in a definition of minimum performance requirements for clinical mastitis detection, which must be met if the detection system should be implemented in a production herd. Both 70 and 80 have been suggested as minimum sensitivity levels, whereas a specificity of 99 is agreed on as minimum level (Mein and Rasmussen, 2008; Rasmussen, 2002). Although no minimum requirements have been defined for detection of other conditions than clinical mastitis, it is a general challenge to reduce the number of false alarms from livestock detection systems (Berckmans, 2014), and it can be argued that the performance requirements defined for clinical mastitis detection, could be used as guidelines for evaluation of livestock detection systems in general.

Only five sensor-based detection models described in the scientific literature, fulfill the defined minimum performance requirements, when comparing to either of the minimum requirement for Se:

- Mol and Ouweltjes (2001) ($Se = 100$, $Sp = 99.5$) used Fuzzy Logic to detect clinical mastitis, but also used the same data set for learning and testing the model
- Liu et al. (2009) ($Se = 100$, $Sp = 100$) validated their B-spline transformed logistic regression model for lameness detection by "leave-one-out" for each of the 260 cows included in the study, which leads to almost identical learning and testing data sets as well
- Kamphuis et al. (2010b) ($Se = 71.4$, $Sp = 99$) detected clinical mastitis and used a decision tree in combination with the *ensemble classifying* methods *bagging* and *boosting* (Witten and Frank, 2005) to obtain the high performances
- Maertens et al. (2011) ($Se = 90$, $Sp = 100$) used linear regression in combination with an unspecified method, and obtained high performances on identifying severely lame (gait score 3) cows correctly, but reported no performance for lower gait scores
- Cornou and Lundbye-Christensen (2011) ($Se =100$, $Sp = 100$) applied a Dynamic Generalized Linear Model (DGLM) to detect sow parturition, but the high performances were to some extent caused by an overfitting of the model, as was recognized by the authors.

The dynamic approaches used by Cornou and Lundbye-Christensen (2011) and Kamphuis et al. (2010b) seem well suited for modeling time series of sensor data. This is further supported by high performances obtained in seven studies, all using Dynamic Linear Model (DLM) with or without postprocessing methods (see Table 1.1).

Despite the high performances obtained in studies using dynamic linear models, too many false alarms are still generated, and the developed models are not suited for implementation in livestock production herds without further customization. A possible explanation for the general difficulties of reaching minimum performance requirements might be that sensitivity and specificity are epidemiological terms designed for evaluation of binary test output. The majority of conditions in livestock production, like clinical mastitis, lameness, oestrus, diarrhea, and fouling, progress over time and are more complex than binary conditions (Friggens et al., 2007, 2010). A precise detection of such conditions will therefore be difficult when using a fixed threshold for generating alarms.

Table 1.1: Sensitivity and specificity obtained by various DLM applications

Paper	Method	Se	Sp	Focus
Cornou et al. (2008)	Univariate DLM	75	95.4	Oestrus
Cornou and Lundbye-Christensen (2010)	Multivariate DLM	96	96	Activity types
Ostersen et al. (2010)	Multiprocess DLM	89.2	96.9	Oestrus
Mol et al. (2013)	DLM	100	95.4	Clinical Mastitis
Jensen et al. (2016)	Multivariate DLM	80	81	Clinical Mastitis
Jensen and Kristensen (2016)	Multivariate DLM	80	88	Diarrhea
Jensen et al. (2017)	Multivariate DLM	80	81	Diarrhea or fouling

In addition to the progressive nature of the conditions, detection performances might be reduced because the modeled parameter contains more information on the animal than what relates directly to the event of interest. Studies indicate that this is true for water consumption of pigs in particular. Thus Jensen et al. (2017) found that water was the one single variable containing most information in the prediction of diarrhea or fouling in finisher pigs, which coincide with the findings by Aparna et al. (2014), in a study predicting the onset of farrowing.

Other studies indicate that pigs' water consumption reflects the true state of the animals more detailed than when the gold standard is directly observed by human. Thus, a study by Andersen et al. (2016) showed that changes in the diurnal drinking pattern of finisher pigs could indicate the presence of stressors in general, rather than specific events. This is also considered in a study by Madsen et al. (2005), who found that changes in the drinking pattern of weaner pigs can be used to predict outbreaks of diarrhea, but also that it may reflect the general wellbeing of the pigs.

If alarms caused by reduced wellbeing in the animals are communicated to the farmer, the warning system would be less event-specific. Unspecific warning systems have a high potential of predicting very early stages of a condition or other stressors amongst the animals. This is of considerably managerial value, but some sorting or prioritization of the unspecific alarms are necessary though. The alarms could be presented to the farmer as risk indicators, as suggested by Friggens et al. (2007, 2010), or they could be prioritized by including non-sensor information in a Naïve Bayesian Network (NBN), as was done with success by Steeneveld et al. (2010).

An alternative alarm-prioritizing approach is to relate the unspecific alarms to a specific area of the herd using a spatial model. This would allow the farmer to include any specific knowledge of the animals in the targeted areas and choose the right intervention timely enough to prevent an outbreak or reduce the consequences of the condition.

Modern Danish production sites for growing pigs are very well suited for such spatial modeling due to the construction of the sites and the managerial routines. The sites are organized in identical sections consisting of a number of identical pens. Pigs are inserted in the sections following an AIAO strategy where pigs of same age are inserted in a section on the same day, and the section is emptied and disinfected before the insertion of a new batch of pigs. This construction of the herd allows it to be modeled as a system consisting of one large unit (the whole herd), which consists of a number of identical subunits (sections), with each subunit consisting of a number of identical sub-subunits (pens), as illustrated in Paper II.

Standardized managerial routines cause a high degree of correlation between pigs in the individual pen as well as between pens within a section, and sections within a herd. Pigs are inserted in the individual pens in a section according to gender or size, and for bio-security reasons pigs from one section do not enter another section in the herd. Both feed mixture and climate control is managed individually for each section according to the age of the pigs it contains. It is, however, the same central feeding system which supplies the whole herd, and both power and main water supply are central as well.

The AIAO strategy described above, constrains the spreading of conditions like diarrhea and fouling between sections. If the conditions do occur, they will spread within a section from one or few pens to more, but it is not given that all pens within a section get affected.

Some points in time during the growth period are known to constitute higher risk of outbreaks than others. Thus outbreaks of diarrhea have shown to be related to shifts in the environment experienced by both weaner and finisher pigs at insertion in the section (Pedersen, 2012) (and personal communication Weber, N.). Routinely shifts in the age-optimized feed mixture two to four weeks after insertion, often cause diarrhea amongst weaners as well (ibid.). In a study by Aarnink et al. (2006), it was found that outbreaks of fouling is closely related to room temperature and body weight. Although pen density and climate related parameters like humidity and draft have some effects (Huynh et al., 2005; Randall et al., 1983), most fouling occur amongst finishers in the end of the growing period, where the increased emission of body heat adds to the environmental temperature (Aarnink et al., 2006; Spoolder et al., 2012).

The development of a spatial detection system for Danish production units for growing pigs, will make it possible to predict events at separate spatial levels within the herd. The system will be able to identify irregularities in a specific pen within a specific section, in a specific section within the herd, or in a pen within the herd. Such area-specific alarms allow the farmer to include any specific knowledge of high risk periods and of the specific animals in the pointed area, and choose the best suited intervention, as discussed in Paper III.

CHAPTER 1 REFERENCES

Aarnink, A. J. A. et al. (2006). „Temperature and body weight affect fouling of pig pens". In: *Journal of Animal Science* 84, pp. 2224–2231. DOI: 10.2527/jas.2005-521.

Andersen, H. M. L., E. Jørgensen, and L. J. Pedersen (2016). „Using Evolutionary Operation technique to evaluate different management initiatives at herd level." In: *Livestock Science* 187, pp. 109–113.

Aparna, U., L. J. Pedersen, and E. Jorgensen (2014). „Hidden phase-type Markov model for the prediction of onset of farrowing for loose-housed sows." In: *Computers and Electronics in Agriculture* 108, pp. 135–147.

Berckmans, D. (2014). „Precision livestock farming technologies for welfare management in intensive livestock systems. (Special Issue: Animal welfare: focusing on the future.)" In: *Revue Scientifique et Technique - Office International des Epizooties* 33.1, pp. 189–196.

Cameron, R. D. A. (2000). „A review of the industialisation of pig production worlwide with particular reference to the asian region". In: *Food and Agriculture Organization of the United Nations*. URL: {http://www.fao.org/ag/againfo/resources/en/publications/agapubs/awi_concept_pig_product.pdf}.

Cavero, D. et al. (2006). „Mastitis detection in dairy cows by application of fuzzy logic." In: *Livestock Science* 105.1/3, pp. 207–213.

Cavero, D. et al. (2007). „Analysing serial data for mastitis detection by means of local regression." In: *Livestock Science* 110.1/2, pp. 101–110.

Cornou, C. and S. Lundbye-Christensen (2010). „Classification of sows' activity types from acceleration patterns using univariate and multivariate models." In: *Computers and Electronics in Agriculture* 72.2, pp. 53–60.

Cornou, C. and S. Lundbye-Christensen (2011). „Modeling of sows diurnal activity pattern and detection of parturition using acceleration measurements." In: *Computers and Electronics in Agriculture* 80, pp. 97–104.

Cornou, C., J. Vinther, and A. R. Kristensen (2008). „Automatic detection of oestrus and health disorders using data from electronic sow feeders." In: *Livestock Science* 118.3, pp. 262–271.

Danish Agriculture and Food Council (2010). *Danish Pig Producers and Animal Welfare*. Tech. rep. Danish Agriculture and Food Council.

Friggens, N. C. et al. (2007). „Estimating degree of mastitis from time-series measurements in milk: a test of a model based on lactate dehydrogenase measurements." In: *Journal of dairy science* 90.12, pp. 5415–5427.

Friggens, N. C., M. C. Codrea, and S. Hojsgaard (2010). „Extracting biologically meaningful features from time-series measurements of individual animals: towards quantitative description of animal status". In: *7th International Workshop on Modelling Nutrient Digestion and Utilisation in Farm Animals*. Ed. by D. Sauvant et al. Wageningen Academic Publishers. ISBN: 978-90-8686-156-9.

Hogeveen, H. et al. (2010). „Sensors and clinical mastitis - the quest for the perfect alert." In: *Sensors* 10.9, pp. 7991–8009.

Huynh, T. T. T. et al. (2005). „Thermal behaviour of growing pigs in response to high temperature and humidity“. In: *Applied Animal Behaviour Science* 91.1. ID: 271155, pp. 1–16. DOI: `http://dx.doi.org/10.1016/j.applanim.2004.10.020`.

Jensen, D. B. and A. R. Kristensen (2016). „Comparison of strategies for combining dynamic linear models with artificial neural networks for detecting diarrhea in slaughter pigs.“ In: *Journal of Animal Science* 94.E-Supplement 5, p. 84.

Jensen, D. B., N. Toft, and A. R. Kristensen (2017). „A multivariate dynamic linear model for early warnings of diarrhea and pen fouling in slaughter pigs.“ In: *Computers and Electronics in Agriculture* 135, pp. 51–62.

Jensen, Dan B., Henk Hogeveen, and Albert De Vries (2016). „Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis“. In: *Journal of Dairy Science* 99.9, pp. 7344 –7361. ISSN: 0022-0302.

Kamphuis, C. et al. (2010a). „Decision-tree induction to detect clinical mastitis with automatic milking.“ In: *Computers and Electronics in Agriculture* 70.1, pp. 60–68.

Kamphuis, C. et al. (2010b). „Detection of clinical mastitis with sensor data from automatic milking systems is improved by using decision-tree induction.“ In: *Journal of dairy science* 93.8, pp. 3616–3627.

Kashiha, M. et al. (2013). „The automatic monitoring of pigs water use by cameras.“ In: *Computers and Electronics in Agriculture* 90, pp. 164–169.

Liu, J. et al. (2009). „Enhancing the prediction accuracy of bovine lameness models through transformations of limb movement variables“. In: *Journal of dairy science* 92.6, pp. 2539–2550.

Madsen, T. N., S. Andersen, and A. R. Kristensen (2005). „Modelling the drinking patterns of young pigs using a state space model.“ In: *Computers and Electronics in Agriculture* 48.1, pp. 39–62.

Maertens, W. et al. (2011). „Development of a real time cow gait tracking and analysing tool to assess lameness using a pressure sensitive walkway: the GAITWISE system.“ In: *Biosystems Engineering* 110.1, pp. 29–39.

Mein, G. A. and M. D. Rasmussen (2008). „Performance evaluation of systems for automated monitoring of udder health: would the real gold standard please stand up?“ In: *Mastitis control: from science to practice. Proceedings of International Conference*. Ed. by T. J. G. M. Lam. The Hague, Netherlands: Mastitis control: from science to practice. Proceedings of International Conference, The Hague, Netherlands; Wageningen Academic Publishers. ISBN: 978-90-8686-085-2.

Mol, R. M. de and W. Ouweltjes (2001). „Detection model for mastitis in cows milked in an automatic milking system.“ In: *Preventive veterinary medicine* 49.1/2, pp. 71–82.

Mol, R. M. de et al. (1997). „Results of a multivariate approach to automated oestrus and mastitis detection“. In: *Livestock Production Science* 48.3, pp. 219–227.

Mol, R. M. de et al. (2013). „Applicability of day-to-day variation in behavior for the automated detection of lameness in dairy cows.“ In: *Journal of dairy science* 96.6, pp. 3703–3712.

Ostersen, T., C. Cornou, and A. R. Kristensen (2010). „Detecting oestrus by monitoring sows’ visits to a boar“. In: *Computers and Electronics in Agriculture* 74.1, pp. 51–58.

Pedersen, Ken S. (2012). „Smaagrisediarre“. In: *Dansk Veterinaer Tidsskrift* 6, pp. 18–22.

Randall, J. M., A. W. Armsby, and J. R. Sharp (1983). „Cooling gradients across pens in a finishing piggery“. In: *Journal of Agricultural Engineering Research* 28.3, pp. 247–259. DOI: http://dx.doi.org/10.1016/0021-8634(83)90073-2.

Rasmussen, M. D. (2002). „Defining acceptable milk quality at time of milking“. In:

Sherlock, R. et al. (2008). „Performance evaluation of systems for automated monitoring of udder health: analytical issues and guidelines“. In: *Mastitis control: from science to practice. Proceedings of International Conference*. The Hague, Netherlands: Mastitis control: from science to practice. Proceedings of International Conference, The Hague, Netherlands; Wageningen Academic Publishers. ISBN: 978-90-8686-085-2.

Sorensen, C. G. et al. (2010). „Conceptual model of a future farm management information system.“ In: *Computers and Electronics in Agriculture* 72.1, pp. 37–47.

Spoolder, Hans A. M. et al. (2012). „Effect of increasing temperature on space requirements of group housed finishing pigs“. In: *Applied Animal Behaviour Science* 138.3, pp. 229–239. DOI: http://dx.doi.org/10.1016/j.applanim.2012.02.010.

Steeneveld, W. et al. (2010). „Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems.“ In: *Journal of dairy science* 93.6, pp. 2559–2568.

Wathes, C. M. et al. (2008). „Is precision livestock farming an engineer’s daydream or nightmare, an animal’s friend or foe, and a farmer’s panacea or pitfall? (Special Issue: Smart Sensors in precision livestock farming.)“ In: *Computers and Electronics in Agriculture* 64.1, pp. 2–10.

Witten, I.H. and E. Frank (2005). *Data mining: practical machine learning tools and techniques*. 2nd edition. Morgan Kaufmann, Elsevier.

2

RESEARCH GOALS

2.1 WORKING HYPOTHESES

Hypothesis I: Drinking patterns between pens within a section and sections within a herd of growing pigs are correlated, and this correlation can be modeled using model parameters defined at different spatial levels.

Hypothesis II: Changes in the drinking patterns of growing pigs are influenced by diarrhea and fouling. By monitoring the water consumption simultaneously in multiple pens and sections, outbreaks of the conditions can be detected in specific areas.

2.2 SPECIFIC AIMS

Paper I: In Paper I, the specific aim is to evaluate the performance of sensor-based detection models in the scientific literature, focusing on alarm-reducing methods. An extensive review of the scientific literature is conducted, and performance, modeling methods, and validation methods of the included models are discussed.

Paper II: Paper II addresses Hypothesis I, and the aim of the paper is to develop a multivariate spatial model, which can model correlations between drinking patterns in pens and sections in any herd of growing pigs. The work presented in Paper II is the first of two steps in the development of a full spatial detection system.

Paper III: Paper III addresses Hypothesis II. The aim of the paper is to evaluate the performance of the detection system, based on the detection model developed in Paper II. This is done through a systematic change of model specifications and performance evaluating settings. An additional aim of Paper III is to present and discuss alarm reducing and prioritizing strategies. The work presented in Paper III is the final step of two in the development of a full spatial detection system.

Paper IV: The aim of Paper IV is to exemplify the alarm reducing potential of the detection system developed in Papers II and III.

3

MATERIALS AND METHODS

Data collection for the research presented in this thesis is conducted in accordance with PigIT alliance aims and decisions. The two included herds are chosen by the PigIT alliance in order to represent different age-groups of pigs, and different managerial resources and routines. Likewise, sensor types and sensor placement within each herd is decided by the alliance, and standard procedures for event registration are described in a PigIT protocol (Lyderik et al., 2016).

The following sections include a description of the materials and modeling methods used for Papers II, III and IV in the present PhD study. Since both materials and methods are described in the papers as well, the following sections serve as an overview, including considerations and motivations for the choice of data processing and modeling methods. All modeling was done using the statistical programming language R (R Core Team, 2014).

3.1 HERD DESCRIPTION

Water data from two Danish herds of growing pigs (weaners 7-30 kg and finishers 30-110 kg) lay ground for the research presented in this thesis. Herd A is a commercial finisher herd, and Herd B is the experimental weaner herd,"Grønhøj", which is owned by the Danish Pig Research Centre[1].

Both herds conduct AIAO management, thus inserting pigs of same age in all pens within a section on the same day, followed by a complete emptying, cleaning and disinfection of the section before new pigs are inserted. Pigs inserted in the same section at the same time are defined as a batch of pigs, and the period they stay in the section is defined as the growth period. The growth period for finisher pigs in Herd A lasts 14 weeks including one week for cleaning, whereas the growth period for weaner pigs in Herd B lasts 8 weeks including four days for cleaning.

Herd A consists of five sections, of which four are included in this study. Herd B consists of four sections in total, and all of those are included in this study. Batches are inserted in subsequent sections in the herds following a production cycle as illustrated in Figure 3.1 from Paper II. This implies that pigs are of different ages across the sections in the herd at any given time. In Herd A the study period was initiated in May 2014 and ended in March 2016, thus monitoring water data from seven batches from each section during the study. Water data from Herd B was monitored from

[1] www.pigresearchcentre.dk

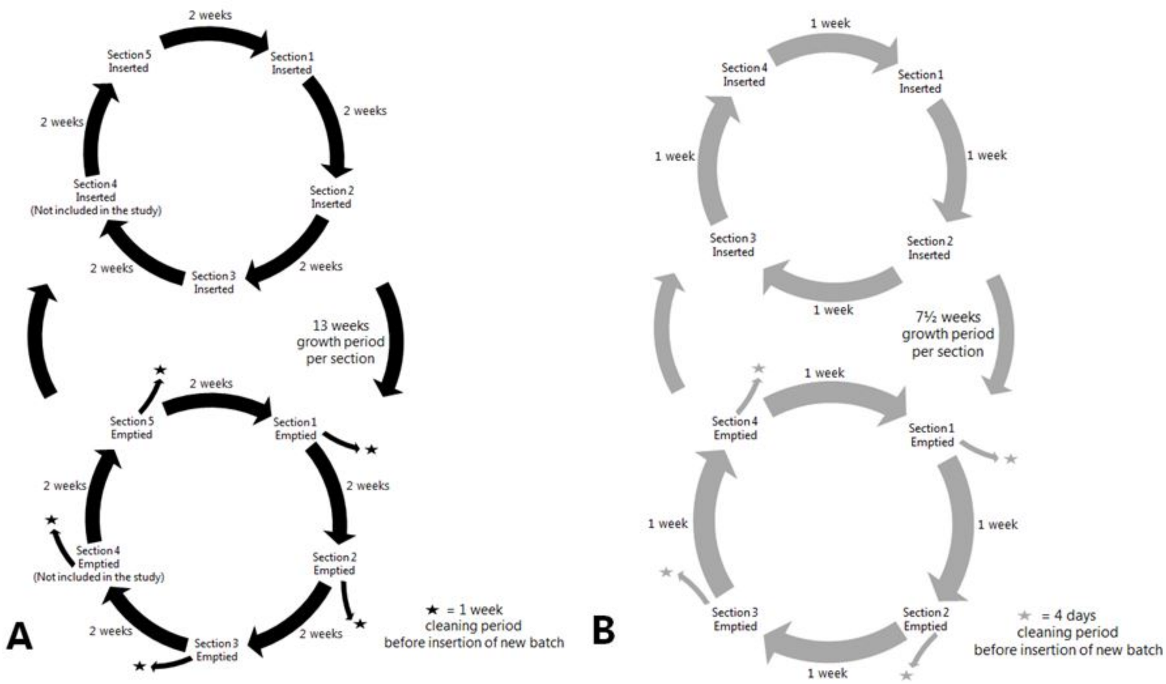


Figure 3.1: Production cycle for a finisher herd (A) and a weaner herd (B) (Figure from Paper II).

October 2014 to December 2016, which included 13 batches from three sections and 14 batches from one section.

A section in Herd A consists of 28 pens each with 18 finisher pigs. Two neighbouring pens share the same feeding trough and the same water pipe. The water consumption of two pairs of neighbouring pens, called double pens, per section is monitored, and both sections and pens within the sections are randomly chosen. A section in Herd B consists of 12 pens each with 15 weaner pigs. All pens have individual feed and water supply, and four pens from each section are included in the study. Both sections and pens are chosen by the Pig Research Centre. The structure of the two herds, and the placement of the pens included in the research, can be seen in Figure 3.2 from Paper II.

3.2 SENSOR DATA

Water data was monitored using photo-electric flow sensors (RS V8189 15mm Diameter Pipe). The sensors measured water flow per millisecond as pulses proportional to the velocity of the water (Anonymous, 2000), and the number of pulses were converted to litres and aggregated per hour, yielding water use in litres per hour.

The sensors were placed on water pipes supplying either double pens in Herd A, or single pens in Herd B (Figure 3.3). In Herd A a total of eight sensors were installed, thus monitoring the water consumption of eight double pens containing a total of 36 finisher pigs each. In Herd B sixteen sensors were installed, each monitoring the water consumption of 15 weaner pigs.

Sensors were calibrated before initiations of new batches, and no sensor data was obtained during cleaning periods between batches. Therefore the cleaning periods are considered planned periods of

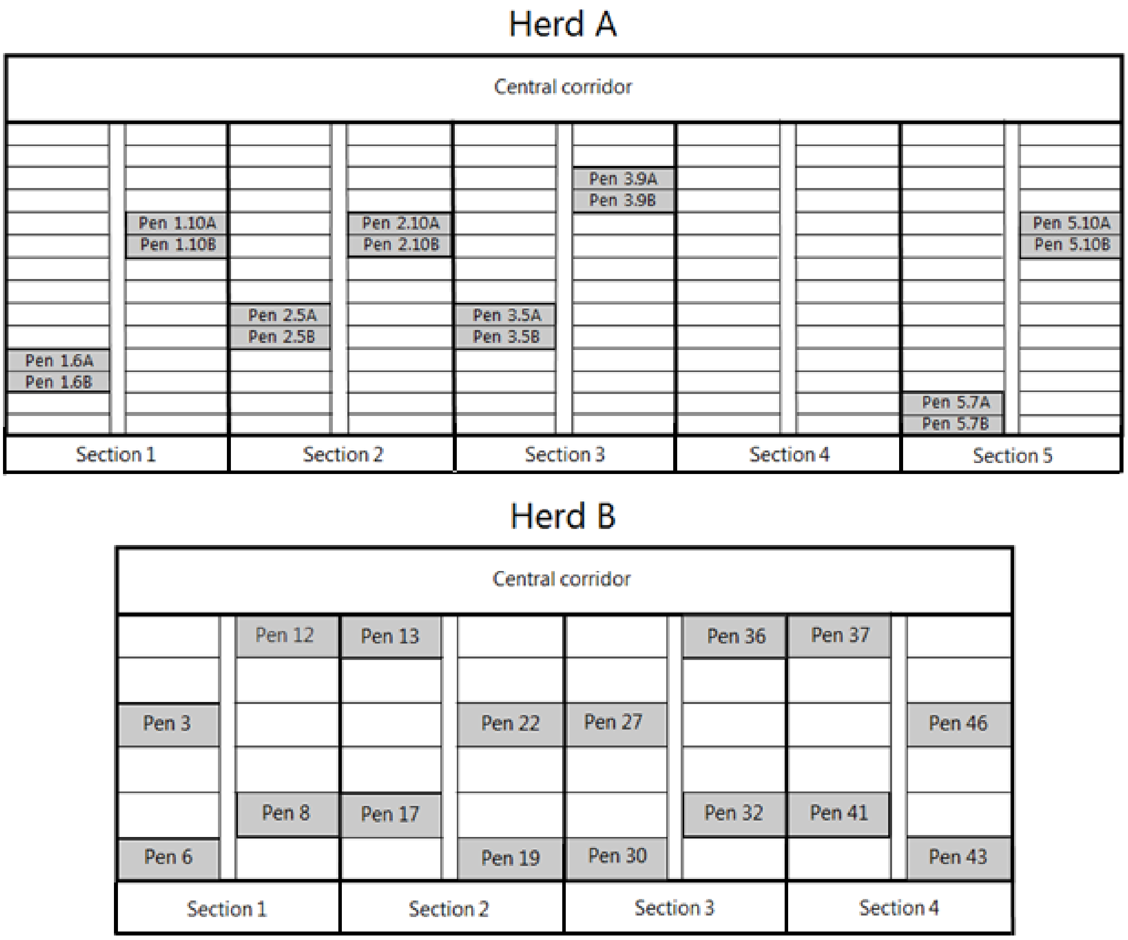


Figure 3.2: Plan of Herd A (top) and Herd B (bottom). Sensors were placed in eight double pens in Herd A (marked with grey) and in sixteen pens in Herd B (marked with grey). (Figure from Paper II).

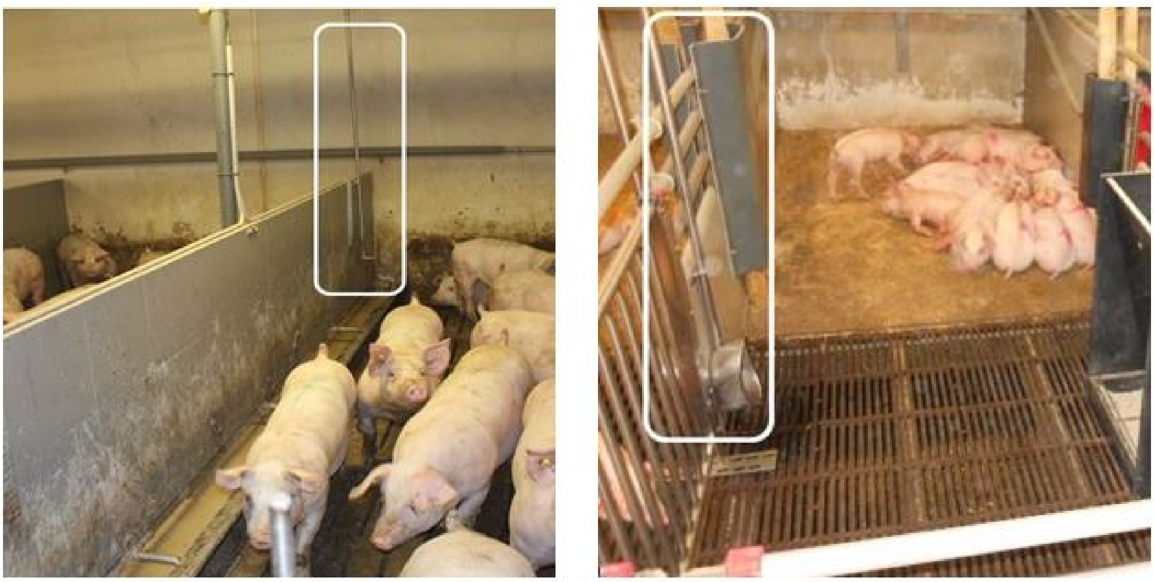


Figure 3.3: One water pipe supplying two neighbouring pens in Herd A (left) and a single pen in Herd B (right). (Figure from Paper II).

missing data, as opposite to unplanned periods of missing data, which are seen in the data sets as well. Some of these unplanned periods are caused by sensor outages, or other technical irregularities in the process of transferring data from a sensor to the central data base. The periods may last from a few hours to several days, although few periods as long as a whole batch are also registered.

In Herd A unplanned periods of missing data occurred almost every night in all pens. Since it is not possible to distinguish between sensor outages and hours without drinking activity (zero observations) in the present data sets, it was suspected that the missing data reflected sleeping pigs. Samples of video recordings from the pens confirmed the suspicion, and therefore missing data of maximum 5 hours length occurring between 10:00 PM and 4:00 AM, as standard, were interpreted as zero observations. In Herd B water consumption was registered in every hour during the night, which may be related to a higher metabolism in smaller animals (weaners) than in larger (finishers). The cause of this night activity is, however, beyond the scope of this research, and will not be discussed further.

Data from each sensor constitute an individual time series, and the full data set for Herd A therefore consists of eight time series, whereas the full data set for Herd B consists of sixteen time series. A full data set begins with the first observation by any sensor in the herd, and ends with the last observation by any sensor in the herd.

The full data sets are split into learning data (4 batches from Herd A and 10 (11) batches for Herd B) used to train the detection model in Paper II, and test data (2 batches from both herds) used to evaluate the performance of the detection model in Paper III. Data from one batch per sensor was omitted between training data and test data for both herds, in order to avoid the risk of having observations from the same pigs occurring in both data subsets.

3.3 EVENTS OF INTEREST

During the study period, the personnel in both herds registered on a daily basis, if either diarrhea or fouling had occurred in any of the monitored pens. These registrations of the two event types constitute the gold standard in the evaluation of the model's predictive performance in Paper III.

At the initiation of the study period, the personnel was given instructions on how to assess the two types of events, and a protocol with detailed descriptions were handed out (Lyderik et al., 2016). In Herd A the assessment routines were calibrated by a trained technician at regular visits, however this was not considered necessary in Herd B since the herd is a research facility. For the rest of the present thesis, no distinction between diarrhea and fouling are made, though. This approach is chosen since the aim of the detection system is to generate area-specific alarms rather than distinguish between different conditions. Registrations of diarrhea and fouling are therefore merged for each herd under the common term "event".

When examining the event registrations from the two herds, it was evident that significant herd-specific differences occurred in the frequencies of registrations. Herd A experienced multiple replacements of the daily caretaker during the study period, and this resulted in inconsistent event registrations including periods with none at all. Therefore all available event registrations were included as gold standard. In Herd B, periods of 14-21 days in a row with positive diarrhea registrations occurred multiple times. None or few treatments were initiated in those periods, though, and conferring with the daily manager confirmed that the personnel's threshold for identifying diarrhea was very low. Thus, for Herd B, the initiation of an intervention (medical treatment of diarrhea or cleaning of pens with fouling) were used as gold standard for performance evaluation.

3.4 MODELING METHODS

Initial evaluation of the drinking patterns from both Herd A and Herd B showed a clear diurnal pattern (see Figure 3.4). Such a diurnal drinking pattern has previously been modeled by Madsen et al. (2005), who used a superpositioned univariate dynamic linear model to model the drinking pattern of a section of weaner pigs. A dynamic linear model is dynamic by nature. It can model fluctuations over time in the underlying mean, which makes it well suited for modeling the evolution in pigs' water consumption over time. Therefore, a multivariate spatial dynamic linear model is chosen as the modeling method in this research, and the model, developed in Paper II, is made on the basis of the work by Madsen et al. (ibid.).

The characteristics of a general multivariate DLM, can be described, following West and Harrison (1999), as:

The observation vector $Y_t = (Y_{1t}, \ldots, Y_{nt})'$, contains the observation at time t for each of the n sensors. Both the relation between Y_t and the underlying parameter vector θ_t at time t, and the evolution of the system over time, are described through an observation equation and a system equation (Equations (3) and (4) respectively):

OBSERVATION EQUATION

$$Y_t = \mathbf{F}_t'\theta_t + \nu_t, \quad \nu_t \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_t), \tag{3}$$

and

SYSTEM EQUATION

$$\theta_t = \mathbf{G}_t'\theta_{t-1} + w_t, \quad w_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t). \tag{4}$$

The overall aim of the DLM is to predict the next observation of the monitored variable by estimating the parameter vectors, $\theta_1, \ldots, \theta_t$, from the observations, $Y_1, \ldots, Y_t$. Every observation is added to the model's prior knowledge of the modeled system, and this dynamic updating enables the model to predict the next observation with increased certainty over time. When a new observation is made, the predicted value and the observed value are compared, and any differences between the predicted and the actual observations are due to the two error terms, ν_t and w_t.

For the modeling of pigs' drinking patterns, this means, as quoted from Paper IV, that "if the pigs follow their normal drinking pattern and drink as much water as expected, the prediction of the next observation is close to perfect, and any prediction error will be small. If, on the other hand, something is causing the pigs to drink more or less than expected, the prediction error will be larger. A systematic change in the normal drinking pattern will generate a sequence of larger prediction errors, and this will lead to an alarm, as will be described later."

3.4.1 *Modeling diurnal patterns*

Madsen et al. (2005) tested how many harmonic waves should be included in order to model the diurnal drinking pattern of pigs. They found, that three waves were sufficient, and in the work presented in this thesis, the same three harmonic waves are used. Therefore, the modeling of the diurnal part of the drinking pattern is done by combining three harmonic waves of lengths 24h, 12h, and 8h as illustrated in Figure 3.4. Each harmonic wave is expressed as cyclic models in a dynamic linear model through the trigonometric *Fourier form representation of seasonality* (West and Harrison, 1999) as:

$$\mathbf{F}_t^h = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{G}_t^h = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}. \tag{5}$$

with $\omega = 2\pi/24$ yielding a wave with a period of 24h, $\omega = 2\pi/12$ a wave with a period of 12h, and $\omega = 2\pi/8$ a wave with a period of 8h. The three harmonic waves are denoted, H1, H2, and H3 respectively.

However, since pigs drink more water as they grow, a trend must be added to the diurnal pattern in order to model the full drinking pattern. A dynamic *linear growth model* models the underlying level of water consumed as well as the increase in the level from time $t-1$ to t. It is described by West and Harrison (ibid.) as:

$$\mathbf{F}_t^l = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{G}_t^l = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{6}$$

3.4.2 *Modeling correlation*

All four elements found by Madsen et al. (2005) always contribute to the pigs' drinking patterns at the same time, but in Paper II it is investigated whether each of the three wave elements, independently of each other, peak at the same time in all pens in a herd, only at the same time in pens within a section, or if there are individual differences in the time of peaking between all pens in a herd.

If a wave peaks at the same time in all pens in the herd, then that element of the full drinking pattern is correlated at herd level (see Table 3.1). This could for example be the wave with 24h wavelength (H1), and this would then indicate that the pigs have their majority of water intake at the

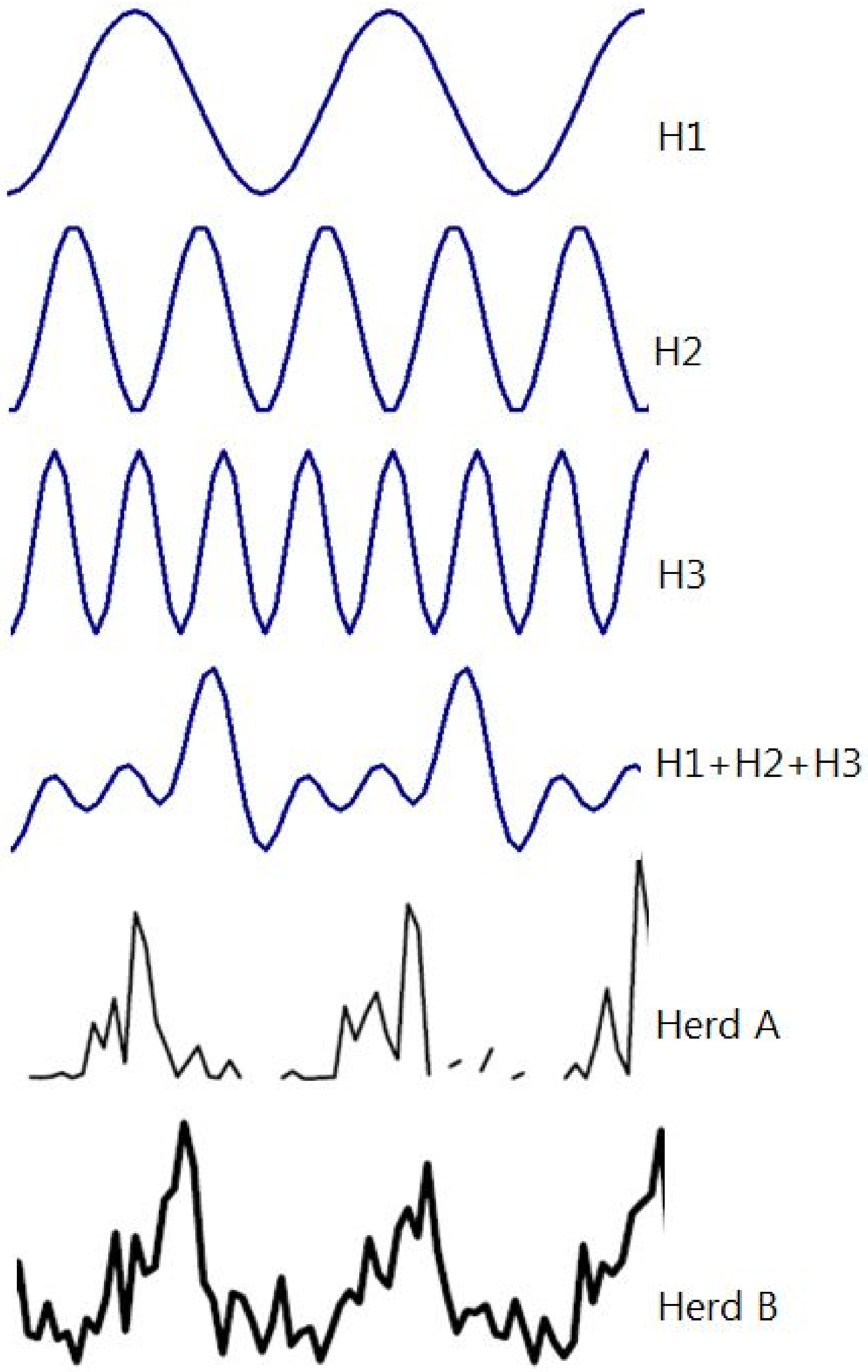


Figure 3.4: From top to bottom: H1 (24h), H2 (12h), H3 (8h), Sum of H1+H2+H3, Diurnal drinking pattern Herd A, Diurnal drinking pattern Herd B

same time every day (for example at noon) in all pens across the herd, and their minimum of water intake 12 hours later (at midnight). Since all pigs in the herd experience night and day at the same time, this correlation would not be unexpected to find.

If, on the other hand, all pigs in one section have their maximum water intake at 10 AM (and minimum at 10 PM), and all pigs in another section have their maximum water intake at 1 PM (and minimum at 1 AM), then this would indicate a correlation in the 24h wave element of the drinking pattern at section level. Since the pigs differ in age and size between sections, this correlation would not be unexpected to find either.

Finally, if differences in the time of day, where the major water intake takes place, are found between all pens in the herd, it would indicate little or no correlation in the 24h wave element of the drinking pattern, and the correlation would then be expressed at pen level. Bearing the AIAO sectionalized structure of a herd in mind, such a correlation would be less expected to be found for the 24h wave. It would, however, not be unlikely for the 12h (H2) or the 8h wave (H3) since this would express differences in the minor peaks of drinking activity between pens.

Table 3.1: Terminology and interpretation of seven model versions applied to the data sets from Herd A and Herd B. The letters denoting the model versions are: H = Herd level, S = Section level, and P = Pen level. In the interpretation, H1 = Harmonic wave of length 24h, H2 = Harmonic wave of length 12h, and H3 = Harmonic wave of length 8h.

Model version	Interpretation
HHH:	The full harmonic pattern evolves identically for all pens in the herd
HSS:	H1 evolves identically for all pens, H2 and H3 evolve identically within each section but differently between sections
HSP:	H1 evolves identically for all pens, H2 evolves identically within sections but differently between sections, H3 evolves differently in each pen
SSS:	The full harmonic pattern evolves identically within each section but differently between sections
SSP:	H1 and H2 evolve identically within sections but differently between sections, H3 evolves differently in each pen
SPP:	H1 evolves identically within sections but differently between sections, H2 and H3 evolve differently in each pen
PPP:	The full harmonic pattern evolves differently in each pen

3.4.3 *Estimating variance components*

For each of the seven model versions, the variance-covariance matrices, $\mathbf{V}_t$ and $\mathbf{W}_t$, are estimated on the learning data. Three observation variance components are defined at herd level, section level, and pen level, respectively, to allow for observational errors to occur in one pen, in all pens in a section, or in all pens in a herd. By combining the three spatially defined variance components, the full variance-covariance matrix, $\mathbf{V}_t$ is expressed. The system variances, $\mathbf{W}_t$, are modeled as a fixed proportion of the posterior variances, $\mathbf{C}_t$, using discount factors. Thus, for each of the three cyclic models as well as for the linear growth model, a discount factor is estimated.

For estimation of the variance components, the Nelder-Mead algorithm in the `optim` function in R (R Core Team, 2014) is used. The algorithm optimizes all parameters with respect to an optimization criterion. In the present case, the parameters are the variance components of the model (three $\mathbf{V}_t$ components and four discount factors used to express $\mathbf{W}_t$), and the optimization criterion is to minimize the Mean Square Error (MSE). When variances are estimated in order to optimize the MSE, there is a risk of *bias-variance tradeoff*, which leads to excessively large variances and small MSEs, or vice versa (Hawkins, 2004; Torgo, 2017; Witten and Frank, 2005).

Variance components can be estimated by other methods than optimization. They can be considered unknown and be estimated for each observation using a Kalman filter, as done by Madsen et al. (2005), or estimated more directly through an *Expectation-Maximization* (EM) algorithm, as done by Jensen et al. (2017) and Bono et al. (2012). In the presented work, the estimation by the EM algorithm was initially attempted, but the iterative algorithm repeatedly failed to converge. This finding was also made by Madsen (2001), and it may be that the EM algorithm is unsuited for estimation in models which include harmonic elements based on the Fourier form representation of seasonality, as mentioned in Paper II. In the presented research, the estimation through optimization is therefore sought investigated, and the outcome of that will be discussed in Section 4.2.4.

3.4.4 *Model fit*

In order to express how well the models predict the next observation in the water data, the Mean Square Error MSE is calculated individually for each model version and each herd. The MSE is defined as $\frac{1}{T}\sum_{t=1}^{T} e_t' e_t$, and expresses the average of the squares of the forecast errors. If a model fits the data well, the differences between the predicted, or forecasted, pattern and the observed pattern will result in smaller errors than for a model, which fits the data less well. Therefore the interpretation of MSE is that a smaller numerical value indicates a better model fit. The MSE can not be compared between herds, but should be compared between model versions applied to the same data set.

3.4.5 *Cusum control chart*

Each modeled time series of sensor observations generates a series of forecast errors, e_t. Monitoring of forecast errors can be used to detect whether a process is in control or out of control, as described in Paper III. Out of control situations may lead to an alarm, which may proceed unwanted events. In Paper III a *two-sided tabular Cusum control chart* is used to monitor the generated forecast errors of each model version. The model versions are evaluated on their ability to detect unwanted events in either a specific pen, a specific section, or in any monitored pen within the herd.

Since the multivariate dynamic linear model generates one series of forecast errors per pen, more forecast errors must be added together in order to express errors at section and herd level. The procedure for preparing the forecast errors to the Cusum is described in details in Paper III, and only a short overview is presented here.

The pen level input to the Cusum consist of the series of forecast errors from the sensor in the corresponding pen (8 in Herd A, 16 in Herd B). The series of forecast errors for a section (4 in both herds) is generated by adding the forecast errors of all sensors within the section at time t. Likewise, the series of forecast errors for the entire herd (1 in both herds) is generated by adding the forecast errors of all sensors within the herd at time t together.

Because pigs drink more as they grow, the numerical values of the forecast errors increase over time as the underlying level of water consumption increases. The growth-related increase must be eliminated if an increase caused by a systematic change is to be identified. Therefore the forecast errors are standardized with respect to the forecast variances, Q_t, before they are monitored. Series of forecast variances for standardization are generated for each spatial level following the method described in Paper III.

Since each series of forecast error and forecast variances is generated for a specific pen, section or herd, it is denoted e_t^u or Q_t^u respectively, where u relates to the specific unit (pen, section, herd).

The following is quoted from Paper III and describes the "applied standardized two-sided Cusum control chart is defined by Montgomery (2013) as:

Since the expected value of e_t^u is 0, the standardized value y_t^u simply becomes

$$y_t^u = \frac{e_t^u}{q_t^u}, \tag{7}$$

where $q_t^u = \sqrt{Q_t^u}$.

Then, the **Upper Cusum** for the unit is the series

$$C_t^{u+} = max[0, y_t^u - k + C_{t-1}^{u+}] \tag{8}$$

and the **Lower Cusum** is the series

$$C_t^{u-} = max[0, -k - y_t^u + C_{t-1}^{u-}]. \tag{9}$$

where k is the *reference value*".

The two-sided Cusum generates cumulated sums of the positive and the negative forecast errors separately over time and plot them as *Upper Cusum* and *Lower Cusum* respectively. If either of the Cusums exceed a defined threshold, h, the monitored process is considered out of control, and an alarm is generated (see Figure 3.5).

Subsequently the generated alarms are categorized as true positive (TP), false positive (FP), true negative (TN), or false negative (FN) based on whether they occur within a chosen time window or not. Three different time windows were defined, each including either three, two, or one day before an event, and no days after (denoted 3/0, 2/0, 1/0 respectively). The time windows are illustrated in Figure 1.1, Section 1.2.

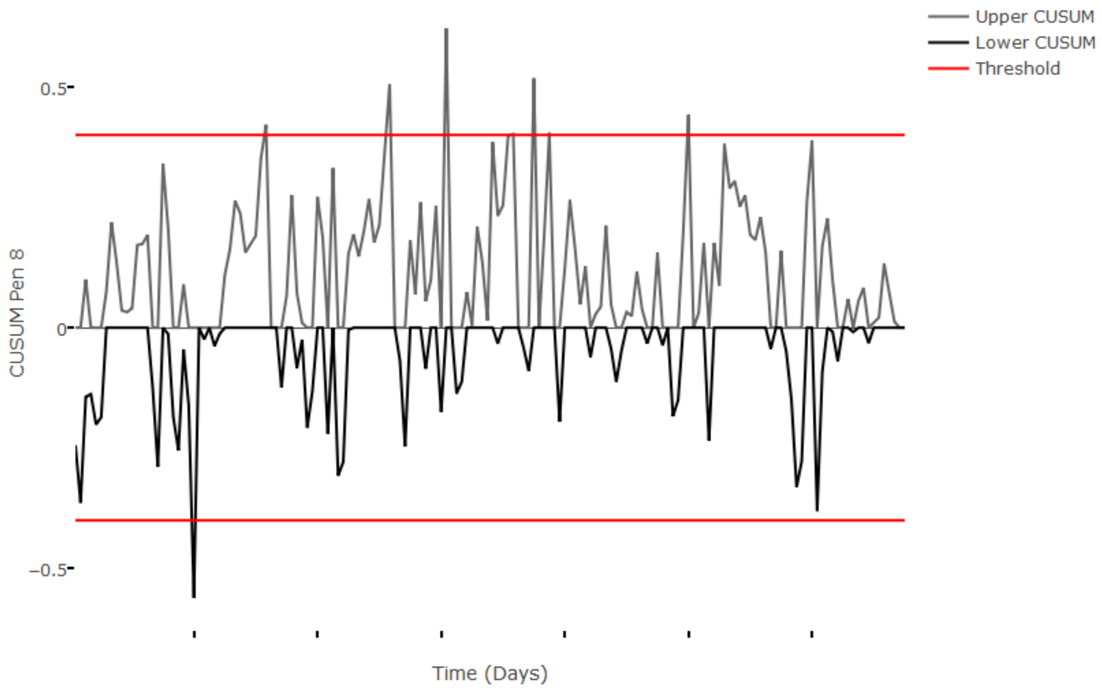


Figure 3.5: Two-sided Cusum control chart. The Upper Cusum reaches or exceeds the threshold seven times, whereas the Lower Cusum exceeds the threshold once. In total eight individual alarms are generated.

The process of the full performance evaluation is described in details in Paper III. A total of $2 \times 7 \times 3 \times 3 = 126$ model combinations are evaluated for their detection performance based on:

- Herd (Herd A, Herd B)
- Model version (HHH, HSP, HSS, SSS, SSP, SPP, PPP)
- Spatial level (Pen, Section, Herd)
- Time Window (3/0, 2/0, 1/0)

For each of the 126 model combinations a Cusum is run with different settings of the two control chart parameters, k and h, and the final performance of the detection system is evaluated independently of a threshold using Receiver Operating Characteristics (ROC) curves and calculating the Area Under the Curve (AUC) for each curve as described in Paper III. If the AUC = 1, then the predictive performance is perfect, so values close to 1 are preferred.

CHAPTER 3 REFERENCES

Anonymous (2000). *Instruction Leaflet Flow sensors - 15 mm Dia. pipe V8189*. URL: http://docs-asia.electrocomponents.com/webdocs/001b/0900766b8001bb47.pdf.

Bono, C., C. Cornou, and A. R. Kristensen (2012). „Dynamic production monitoring in pig herds I: modeling and monitoring litter size at herd and sow level.“ In: *Livestock Science* 149.3, pp. 289–300.

Hawkins, Douglas M. (2004). „The Problem of Overfitting“. In: *Journal of chemical information and computer sciences* 44.1. doi: 10.1021/ci0342472 pmid: pp. 1–12.

Jensen, D. B., N. Toft, and A. R. Kristensen (2017). „A multivariate dynamic linear model for early warnings of diarrhea and pen fouling in slaughter pigs.“ In: *Computers and Electronics in Agriculture* 135, pp. 51–62.

Lyderik, K.K. et al. (2016). *Protocol for daily observation in the herds of the PigIT project. PigIT report no 8*. Tech. rep.

Madsen, T. N. (2001). „Tools for monitoring growing pigs“. PhD thesis. The Royal Veterinary and Agricultural University.

Madsen, T. N., S. Andersen, and A. R. Kristensen (2005). „Modelling the drinking patterns of young pigs using a state space model.“ In: *Computers and Electronics in Agriculture* 48.1, pp. 39–62.

Montgomery, Douglas C. (2013). *Statistical Quality Control A Modern Introduction*. Seventh. John Wiley & Sons, Inc.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Torgo, Luís (2017). *Data Mining with R Learning with Case Studies*. Chapman & Hall.

West, Mike and Jeff Harrison (1999). *Bayesian Forecasting and Dynamic Models*. 2nd. Springer.

Witten, I.H. and E. Frank (2005). *Data mining: practical machine learning tools and techniques*. 2nd edition. Morgan Kaufmann, Elsevier.

4

FINDINGS AND DISCUSSION

In this chapter, the overall findings from the four original papers will be presented and discussed. More in-depth and detailed results are found in the papers in Chapters 5, 6, 7, and 8.

4.1 PAPER I: FOCUS NEEDED ON ALARM REDUCING METHODS

Two primary findings are made in Paper I. The first is that further focus on alarm reducing methods is needed, and the second is that none of the evaluated models are suitable for implementation in commercial herds based on the three evaluation criteria.

Paper I is a literature review, and the initial aim was to evaluate different alarm prioritizing or alarm reducing methods for detection models in livestock production. However, an extensive literature search soon revealed that only three scientific papers focused specifically on this area. Therefore the focus of the review was expanded to include a performance evaluation of sensor-based detection models within the field of livestock production.

The literature search showed that a variety of modeling methods have been applied to different types of livestock related sensor data for the past twenty years. Therefore, a total of 34 included papers were sorted into three groups based on the complexity of their methodology and their focus on prioritizing alarms. All described detection models were evaluated with respect to minimum performance demands (sensitivity = 80 and specificity = 99) defined in the scientific literature. Subsequently the models were evaluated with respect to their study design, which should reflect field conditions, and a maximum length of 48 hours time window, if one was applied.

4.1.1 *Alarm reducing methods*

As mentioned above, only three of the evaluated papers focus specifically on alarm reducing methods. In the first paper, Mol and Woldt (2001) use *Fuzzy logic* to reduce the number of false alarms. The method is, however, developed for quantifying linguistic values like “a little”, “to some extent”, or “very much”, and is not well suited for sensor-based data sets of numeric variables (Klir and Folger, 1988).

In a second paper, Aparna et al. (2014) model the successive behavioural patterns undergone by sows prior to farrowing. They then use the *Hidden phase-type Markov* method to predict the precise onset of the farrowing. This approach is well suited for conditions known to happen, such as the farrowing, and with well-defined behavioural stages preceding the event. The general situation for detection models in livestock production is, however, an uncertainty of whether the condition will occur or not. Additionally, knowledge on a well defined behavioural pattern prior to a condition occurring is generally lacking.

The third paper is by Steeneveld et al. (2010a) and introduces *Naïve Bayesian Network* (NBN) as a successful method for discriminating between true positive TP and false positive FP alarms in the detection of clinical mastitis. By combining information from the automatic milking system with non-sensor cow-specific information, the probability of an alarm being true or false is calculated, and the number of TP alarms is reduced significantly.

Of the three methods, NBN is found to be the better suited for sensor based models detecting unforeseen events in livestock production. This is supported in the findings of Jensen et al. (2016), who recently combined sensor and non-sensor information using a dynamic linear model with a *Naïve Bayesian Classifier* (NBC) for clinical mastitis detection as well.

Kamphuis et al. (2010b) used a *decision tree* combined with the ensemble classifying method *bagging* to improve the performance for detecting clinical mastitis. Ensemble classifying methods are machine learning methods, which combine the output of different models, or of the same model trained on different data sets, to increase the predictive performance over a single model as described by Witten and Frank (2005). Retrospectively, the methods used by Kamphuis et al. (2010b) should therefore have been categorized as an alarm-reducing method as well. Machine learning methods are well suited for analyzing large amounts of data (Witten and Frank, 2005), but they are not yet widely used for detecting conditions in livestock production. However, if the amount of sensor-based data in livestock production increases as expected (Berckmans, 2014; Sorensen et al., 2010), both decision trees and ensemble classifying methods are likely to be increasingly relevant for further research.

4.1.2 *Performance evaluation*

In Paper I, the sensor-based detection models for livestock production, presented in 34 scientific papers published between 1995 and 2015, are evaluated for their suitability for implementation in commercial herds. The included models were evaluated on three *implementation criteria* initially defined by Hogeveen et al. (2010) in a review of 16 models for detection of clinical mastitis. These criteria are **a)** a sensitivity (Se) of minimum 80% and a specificity (Sp) of minimum 99%, **b)** a time window of maximum 48 hours, and **c)** the studies had to be conducted under conditions as similar to practical field conditions as possible. In order to be evaluated in Paper I, the model input also had to be obtained directly from sensors, and not through laboratory analyzes, as was for example done by (Chagunda et al., 2006b; Friggens et al., 2007) who modeled the enzyme lactate dehydrogenase in cow milk.

The evaluation in Paper I show that none of the included detection models, developed over a twenty years period, were suited for implementation in commercial livestock production in their current form. Some of the poorer performances may reflect that focus was on an initial investigation of new modeling methods (Freson et al., 1998; Kamphuis et al., 2010a), but it is still surprising to find that fulfilling the implementation criteria appear so difficult.

Two possible explanations may be lying the ground for the findings in Paper I. The first is that the implementation criteria developed for mastitis detecting models may be too restrictive for models focusing on other conditions. The second is that the evaluation methods, which are generally applied to measure model performances, may not apply for livestock production data. Thus, the use of imprecise gold standards in combination with sensitivity and specificity has been questioned by Chagunda et al. (e.g. 2006a) and Friggens et al. (2007, 2010), and in the following the use of these performance measurements will be discussed further.

4.1.3 *Evaluation criteria*

The first of the three implementation criteria defines the minimum requirements for the detection performance. As described in Chapter 1, the sensitivity reflects the model's ability to identify animals with the condition correctly, and the specificity reflects the model's ability to identify animals, who do not have the condition, correctly. Most animals in livestock production are healthy and sound, which means that the conditions sought detected by detection models generally are rare. Thus, the minimum required specificity of 99% is considered necessary for reducing false alarms from livestock detection systems in general.

The minimum requirement for sensitivity (80%) used by Hogeveen et al. (2010) reflects that undetected cases of mastitis may lead to significant consequences. An undetected cow with clinical mastitis constitutes a risk of infected milk getting in the bulk tank. The consequence of that is, that the full tank of milk has to be discarded, which implies significant economic consequences for the farmer (Rasmussen, 2002). However, not all undetected conditions have equally important consequences in livestock production, and as high a sensitivity may not be required for detecting them as for mastitis.

When considering a condition like lameness, the economic consequences of a detection model missing a lame animal are smaller than for a missed case of mastitis. A (mildly) lame animal is not as high yielding as a sound animal (Garcia et al., 2014), but otherwise the production as a whole is largely unaffected. Within lameness evaluation, different degrees of lameness are defined, and models for lameness detection generally aim to identify early stages of lameness to allow for early interventions (Garcia et al., 2014; Maertens et al., 2011; Mol et al., 2013). However, lameness in general progresses slower than clinical mastitis, and if a mildly lame cow is missed by the detection system, there will be more chances for it to be detected before neither welfare nor productivity is significantly affected.

A minimum sensitivity of 80% for mastitis detection may therefore be more restrictive than what is required for other conditions, and even within the field of mastitis detection a sensitivity of 70% has been suggested (Mein and Rasmussen, 2008). The higher performance criterion is, however, still used in the evaluation of all models in Paper I. This was decided because a sensitivity of 80% is found to be obtainable by human observation in a study by Quimby et al. (2001), and the consideration was that a detection system should be able to match that standard.

The second implementation criterion states that studies should be conducted under conditions as close to field conditions as possible. This criterion is highly relevant if the study aims to develop a model for implementation in commercial herds. It is less significant if the study is focused on testing new methods for data modeling, though, but the aim of Paper I is to evaluate the models for their implementation suitability, which is why the second criterion is applied to all models disregarding the condition in focus.

The third implementation criterion defines a maximum detection window, or time window, of 48 hours prior to the occurrence of the condition. A time window is a defined period associated with the occurrence of a condition, and any alarm generated within that time window are considered true, detecting the condition correctly (see Figure 1.1, Chapter 1). The length of the optimal time window is related to the specific condition, and may be longer for a slower progressing condition, like lameness, than for a faster progressing condition like mastitis. However, in the everyday management of a commercial herd, too long time windows are of little managerial value since the alarm indicates the occurrence of condition with less precision. The 48 hour time window is therefore considered suited for evaluation of models detecting all types of conditions in Paper I.

4.1.4 *The use of performance measurements*

When evaluating the models in Paper I, it was the lack of fulfilling the performance criterion alone, which caused the majority of models to be categorized as unsuited for implementation in commercial herds. In the development process of livestock detection models, it is, however, very useful to measure the sensitivity and specificity of different model versions. A performance comparison of the same model run on different data sets, or different models run on the same data set, can help points the modeler towards the better choice of modeling parameters and methods. This is widely done in studies evaluated in Paper I Cavero et al. (e.g. 2006), Cornou et al. (2008), Kamphuis et al. (2008), and Mol and Ouweltjes (2001) as well as Papers II and III in the present thesis.

However, using sensitivity and specificity to evaluate whether a model is suitable for implementation may not be reflecting the model's true potential after all. A lower performance can be founded in an imprecise *gold standard* registration, or it can be founded in the modeled parameter reflecting the wellbeing of the animals, or even subclinical stages preceding the outbreak of a disease. Such a detection model could contain a lot of information on the animals, and may therefore have significant managerial value. A grading or sorting of the information is, however, crucial for the managerial value to be realised. For the rest of this section, the challenges of gold standard registrations and subclinical stages of diseases is sought elaborated.

4.1.5 *Gold standard challenges*

As briefly mentioned, most unwanted conditions, such as diseases, in livestock production are progressive, which means that all stages can not be correctly identified using a fixed threshold (Friggens et al., 2007, 2010). Furthermore, sensitivity and specificity ares calculated on the basis of a categorization of the generated alarms as true and false by comparison to the *gold standard*, as described in Section 1.2. Hence a true alarm is associated with a registered event, and a false alarm is not.

The registration of the gold standard is usually conducted by human observation, often the daily caretakers in the herds. The presence or absence of the event is registered on the basis of a defined threshold or a description of what the event should look like, when present. For lameness detection, there is no consensus of how to assess different degrees of lameness, which has resulted in more than twenty different lameness scoring scales (Tello et al., 2011). For mastitis detection, a consensus for gold standard definitions has been proposed multiple times (Mein and Rasmussen, 2008; Rasmussen, 2002, 2005), but it has not been reached.

4.1.6 *Subclinical stages*

However, neither consensus nor clear definitions of the gold standard guarantee a true reflection of the state of the animal. The natural subjectivity embedded in human observation will always have an impact, but sometimes it may not even be possible to assess the gold standard by observing the animal. For diseases like diarrhea, clinical mastitis, and laminitis (causing lameness), subclinical stages may precede the clinical stage (Maatje et al., 1997; Somers et al., 2003; Weber et al., 2015). A disease is clinical when it is observable, which means that the subclinical stage is unobservable in a direct assessment of the animal. In a study by Weber et al. (2015), one-third of the weaner pigs, who were assessed healthy by the personnel, actually did suffer from subclinical diarrhea when faecal samples were analyzed.

An animal will always be affected to some extent by the subclinical stage of a disease, and if the modeled parameter in a detection model reflects the true state of the animal, an alarm will be generated. Such an alarm will be classified as a false alarm when compared to the observed gold standard, and reduce the performance measurement of the model. Even though subclinical stages of diseases are not sought detected in the papers included in Paper I, the possible effects of reduced model performance are addressed (Hertem et al., 2014; Kamphuis et al., 2010a, 2011).

In Section 1.2 it was mentioned how changes in the general wellbeing of growing pigs may be reflected in their drinking patterns (Andersen et al., 2016; Madsen et al., 2005). The same is not described for other modeled parameters or other groups of animals, but is is not unlikely, that changes in the general wellbeing affect the outcome of a detection system. If alarms are generated due to changes in the general wellbeing of the animals, they too will be categorized as false alarms.

4.1.7 *Future focus*

A clear distinction between reduced wellbeing and subclinical conditions may not be possible, and such a distinction may not represent a managerial value either. It would, however, be of managerial value to get a warning timely enough to prevent an outbreak of the condition. Both reduced wellbeing and subclinical conditions can progress to unwanted conditions (Andersen et al., 2016; Kamphuis et al., 2011), and therefore alarms generated for the reasons presented above, all contain information on the livestock animals. If the alarms are categorized as false, valuable information may be lost. Therefore, the future challenge is to include the alarms, but in a prioritized structure, or presented in a conceivable form to the farmer. Hereby the total number of alarms communicated will be reduced, while the value of the information is preserved.

Alternative approaches to alarm handling have been suggested or applied. Friggens et al. (2007, 2010) suggested the use of a risk index for grading alarms, Steeneveld et al. (2010b) included non-sensor information in an NBN in order to reduce the number of false alarms, and Kamphuis et al. (2010b) improved the model performance significantly by applying an ensemble classifier. However, a spatial approach has not yet been investigated as an alarm reducing method, which is why this is developed, evaluated and explored in Papers II, III, and IV.

4.2 PAPER II: CORRELATIONS CAN BE MODELED SPATIALLY

In Paper II, I addressed Hypothesis I “Drinking patterns between pens within a section and sections within a herd of growing pigs are correlated, and this correlation can be modeled using model parameters defined at different spatial levels.”

Multiple water sensors were installed in multiple pens across a finisher herd (Herd A) and a weaner herd (Herd B) as described in Section 3.2. Simultaneous monitoring of the drinking patterns in more pens, allowed for interactions, or correlations, between the patterns to be identified. For this a spatial dynamic linear model was developed, and the degree of correlation was sought modeled separately for each of the two herds.

4.2.1 *Correlations in drinking patterns - an introduction*

As described in Section 3.4.2, the full drinking pattern consist of four elements. The fourth element, the underlying level, describe the amount of water consumed by the pigs in a pen over time. Pigs drink more as they grow, which means that the underlying level increases over time as well. Since the sections in the herd each are filled with pigs of same age and size at different times (see Figure 3.1, Section 3.4), the underlying level is assumed to evolve identically for all pens within a section but differently between sections. The model, which described the underlying level, was therefore defined at section level in all model versions.

In both Paper II and Paper III, each model version was denoted by three letters. The first letter responds to the 24h harmonic wave (H1), the second letter responds to the 12h harmonic wave (H2), and the third letter responds to the 8h harmonic wave (H3). The letter "H" reflects a correlation at herd level, whereas the letters "S" and "P" reflects correlations at section and pen level respectively. Thus, the model version, which defines a herd level correlation between all three waves, is denoted "HHH", and the model version defining H1 at herd level, H2 at section level, and H3 at pen level is denoted "HSP" (see Table 3.1). Generally speaking, the more pens a wave describes, the higher the degree of correlation. Therefore, the HHH model version expresses the highest degree of correlation, whereas the PPP model version expresses the lowest.

4.2.2 *Findings Herd A*

As seen in Table 4.1, the SSS model version obtained the best fit (MSE = 13.850) of the seven model versions in Herd A. This model version defined the full drinking pattern at section level and hereby indicated that pigs within the same section were more similar to each other than to pigs of different ages and sizes in other sections. Thus, the sectionalized AIAO production, which characterize herds for growing pigs in Denmark, was reflected in the pigs' drinking patterns in Herd A, as initially expected.

Further results for Herd A (see Table 4.1) showed that model versions, which include minimum one wave parameter at herd level (HHH, HSS, HSP), fitted data the worst. This indicated that the drinking patterns were too different between sections in Herd A to be characterized by the same parameter. Since all sections were placed within the same building and the same central water supply lead to all pens, there has to be some correlation between sections in the herd as well. However, the fits of the three model versions, which included herd level waves, indicated that this correlation was not very strong.

Model versions which included wave parameters at pen and section level (SSP, SPP, PPP) fitted the data better than any version with a herd level parameter, but worse than the SSS model version. This indicated that some pen differences were apparent in the data from Herd A, but these differences were not larger than the section specific similarities.

Since Herd A is a commercial finisher herd, and no major managerial routines were altered for the sake of the present study, it is likely that we can find a similar correlation between pens within sections to be found in other commercial finisher pens as well.

4.2.3 *Findings Herd B*

For Herd B, the fit of the model versions got poorer as the degree of correlation increased (see Table 4.1). Thus, the best fitting model version was PPP (MSE = 1.466), whereas the worst fitting was HHH (MSE = 1.750). This result indicated, that the drinking patterns in the individual pens, both within a section and across the herd, differed too much to be characterized by a common parameter.

Table 4.1: MSE (mean square error) for seven model versions for Herd A and Herd B (test data). The lowest numerical value of the MSE indicate the best model fit. The MSEs are herd-specific and can not be compared between herds. Notations: H1 = Cyclic model of length 24, H2 = Cyclic model of length 12, H3 = Cyclic model of length 8. H = Herd level, S = Section level, P = Pen level

Model Version	MSE	
H1 H2 H3	Herd A	Herd B
HHH	15.687	1.750
HSS	14.535	1.727
HSP	14.612	1.712
SSS	13.850	1.621
SSP	13.976	1.559
SPP	13.946	1.556
PPP	13.924	1.466

This apparent lack of correlation structure between the drinking patterns in Herd B did not match the initial expectations of a section level correlation structure. The correlation in the evolution of the drinking patterns over time, is still expressed by the system variance ($\mathbf{W}_t$) though, and it may be sufficient to describe any interactions in the system. However, since Herd B is a research facility, there are some factors, which were expected to reduce the difference between pens within a section, and hereby increase their correlation as compared to a commercial weaner herd. These factors are **a)** the pigs are more uniform, considering weight and condition, at insertion (unpublished data), **b)** there are more managerial resources available, and **c)** the production environment, including feed management, is highly controlled. Such an increased correlation between pens in a section was, however, not recognized by any of the seven model versions.

Although the results from Herd B were unexpected, explanations can be found in the complexity of the model, the estimation of the variance components, and a high degree of random noise in the data from Herd B. These aspects all point towards an overfitting of the learning data, which will be discussed in the following section.

4.2.4 *Overfitting*

If a model is so flexible that it adjusts to any irregularities, or random noise, in the data set, it is said to be overfitting the data (Hawkins, 2004; Torgo, 2017). Since random noise by nature is random, the noise will be different in all pens, and by adjusting to the random noise, an overfitting model will fail to recognize a general underlying pattern in the data.

In Herd B, a high degree of random noise is present in all pens, which is expressed by drinking activity throughout the night. In all pens, one or two weaner pigs get up at some point during every hour to eat and drink a little, while the rest of the pen sleeps (validated by samples of video recordings). Since a pen in Herd B contains 15 weaner pigs, such night activity by a few pigs constitutes a relatively large degree of random noise. It is therefore likely that the model versions with wave parameters defined at pen level, adjust to the random noise in each pen, and fail to recognize any correlation.

A model can be overfitting for more causes. One is, if the modeled variables are highly correlated. Another is, if the model consists of more parameters than what is needed to describe the pattern in the data. A third is if the estimated variance components are very high as a consequence of a bias-variance tradeoff (Hawkins, 2004; Torgo, 2017; Witten and Frank, 2005).

As described above, the variables, or drinking patterns, are likely to be highly correlated in Herd B. All pigs within a section are very uniform at insertion. In addition to that, pigs are inserted in subsequent sections with one week in between. Of this follows that there is a little weight span between the youngest and oldest pigs across the herd, which ought to support a degree of correlation in drinking patterns at herd level as well. Correlated data, which is the first possible cause of overfitting, is therefore present in Herd B.

The second possible cause for overfitting is excess model complexity. And the spatial dynamic linear model described in Paper II is fairly complex. It includes three variance components (pen level, section level, herd level) for the full observation variance-covariance matrix, $\mathbf{V}_t$, and four discount factors for the full system variance-covariance matrix, $\mathbf{W}_t$ (see Paper II for further description). The observational variances depend on rather constant cause of errors included in the observed data. For a pen, this could be a leaking water bowl, or it could be measurement inaccuracies in the water sensor. For a section this could be a leak or a cloak in the water pipe supplying the section, and for the herd it could be a degree of failure in the central water supply.

However, the influence of observation errors at herd level is rather insignificant in both Herd A and Herd B (see Table 4.2), and a removal of the herd level variance in future work will reduce the complexity of the model without information loss, whereby the risk of overfitting is reduced as well. Although Table 4.2 also shows that the contribution of the section level variance component is very small for Herd B, it is significantly for Herd A, and should therefore not be removed.

The third possible cause for overfitting is a bias-variance tradeoff, which is seen as an increase in the variance estimates when the MSE is sought minimized (Torgo, 2017), as is done in Paper II. High variance estimates do not express the true variance, but the values optimizing the MSE. They do, however, increase the flexibility of the model and enables it to adjust to random noise, as described for Herd B. Variance estimates as high as 176 litres^2/hour are found for Herd A, whereas the estimates for Herd B are even higher with values as high as 6474 litres^2/hour (results in Paper II).

4.2.5 *Conclusion Paper II*

For Herd A, the results of Paper II indicated a correlation in drinking patterns between pens within a section, as initially expected. It hereby addresses Hypothesis I, as stated in Chapter 2, and confirms it.

Table 4.2: Average contributions in % from observation variance components at different spatial levels to the full observation variance. The contributions are calculated from the sum of variance estimates within a level across the seven model versions.

Level	Herd A	Herd B
Herd	2.32	0.13
Section	50.16	1.25
Pen	47.52	98.62

For Herd B, the overfitting of data was likely to cause the PPP model version to obtain the best fit, although it may not reflect the true correlation structure in the herd. Different correlation structures were, however, reflected through other model versions. Although the model should be reduced in complexity and run on data from a commercial weaner herd in order to clarify the degree of overfitting, Hypothesis I is confirmed for Herd B as well.

4.2.6 *Future focus*

The better fitting model versions are not necessarily the best predicting model versions, though. In other words, the ability of a model to detect unwanted events in specific areas of a herd, can not be concluded on the basis of the model fit. A high fitting model may adjust too well to irregularities and fail to recognize changes preceding unwanted events. It will therefore be a better choice to build a detection system on a model with less flexibility. A less flexible model will tend to follow the general pattern, be less adaptive to changes in the pattern, and therefore be more prone to detect them.

The model developed in Paper II constitutes the initial step of two in the development of a full spatial detection system, and in Paper III the seven model versions will be applied to a Cusum control chart, and the detection performances will be evaluated.

4.3 PAPERS III AND IV: AREA-SPECIFIC ALARMS

In this section, the findings of Papers III and IV will be presented and discussed. Paper III evaluates the detection accuracy of the seven model versions defined in Paper II with regard to their managerial value, and discusses both an alarm-reducing and an alarm-prioritizing strategy as well. In Paper IV a reduced number of alarms for a section is compared to the total number of individual alarms for pens within the section. This is exemplified on a data sample from each of the two herds, Herd A and Herd B.

In Paper III, the second working hypothesis is addressed. The hypothesis is, as stated in Chapter 2: “Changes in the drinking patterns of growing pigs are influenced by diarrhea and fouling. By monitoring the water consumption simultaneously in multiple pens and sections, outbreaks of the conditions can be detected in specific areas.”

The overall finding in Paper III was, that area-specific alarms for either diarrhea or fouling could be generated in a herd of growing pigs based on changes in their drinking patterns. The results showed that the HHH model version, which defines the strongest degree of correlation for the full drinking pattern, was the best suited version for detection of unwanted events in both Herd A and Herd B. This overall finding hereby confirms Hypothesis II.

The overall finding in Paper IV was, that 6 individual pen alarms from pens within the same section in Herd A could be reduced to 4 section alarms, whereas 42 individual pen alarms were reduced to 8 section alarms in Herd B.

4.3.1 *Performance evaluation - an introduction*

Each of the seven model versions, defined in Paper II, were all evaluated for their ability to generate area-specific alarms given different lengths of time windows. In this context, "area-specific alarms" are alarms, which can detect the occurrence of either diarrhea or fouling at any of the three spatial levels pen level (in a specific pen), section level (in a specific section), or herd level (in any of the pens in the herd). By nature, the pen levels describe a more specific area than section levels, which again are more specific than the herd level.

As described in Section 1.2, time windows are often defined relative to an event when the performance is evaluated. If a time window includes a period after the event, then the alarm may occur after the event as well, and it will still be classified as a true alarm. Such an alarm has, however, little managerial value, and therefore all time windows in Paper III are defined to include days before the event and the actual day of the event as follows:

- Time window 3/0 includes three days before the event and zero days after
- Time window 2/0 includes two days before the event and zero days after
- Time window 1/0 includes one day before the event and zero days after

The lengths of the time windows were chosen in order to reflect settings that would have a value in the everyday management of a herd of growing pigs. It was taken into consideration that longer time windows allow for less precise timing of managerial interventions, and therefore may encourage the manager to trust the daily visual assessments of the animals to a greater extent than the information from the detection system.

Although the defined time windows in this study are relatively short, they will be prolonged if subsequent events occur before the time window of a previous event has passed. This will lead to overlap of the time windows as illustrated in Figure 1.1, Section 1.2. Prolonged time windows have significant impact on the performance evaluation of events at the different spatial levels as will be discussed later.

When evaluating the performance at pen level, only the events occurring in the specific pen are used as gold standard. However, when evaluating the performance at section level, all days with events registered in any pen within the section are included in the gold standard. The same applies for the evaluation of the herd level performance, which implies that all events registered at any time in any pen in the herd constitute the gold standard.

In the present study, the combination of the longest time window (3/0) and events at herd level, caused overlap of multiple subsequent time windows. As a consequence of that, very high detection performances for both Herd A (AUC = 0.9358, longest time window = 20 days) and Herd B (AUC = 0.9842, longest time window 47 days) were found. Although these performance measurements are close to perfect, the managerial value of such long time windows is low, as is the managerial value of alarms at herd level regardless of time window settings. Based on these findings, these two model settings (herd level alarms and 3/0 time window) should therefore not be investigated further in future studies.

4.3.2 *Findings Herd A*

In Herd A, the HHH model version, which defines correlation in the full drinking pattern between all pens in the herd (see Table 3.1), obtained the overall highest performance of all model versions (results in Paper III). Since the HHH model version obtained the poorest fit in Paper II (see Table 4.1), this finding indicates that the least flexible model version is better at making a clear distinction between general drinking patterns and systematic changes preceding unwanted events in a pen, in a section, and in the herd.

The detection performances (AUC) for the HHH model version in the three spatial levels given all three lengths of time windows are presented in Table 4.3. The results show that the performance measurements increase as the spatial level gets more general, and the time windows get longer. However, all model versions with herd level settings, and those using time window 3/0, constitute low managerial value, as mentioned above, and will not be discussed further here.

The results in Table 4.3 also show that the detection performances are almost identical for pen level and section level given the same length of time window for model version HHH. This indicates that the model detects an upcoming event with the same accuracy whether the alarm is generated for a specific pen or for a specific section in the herd. The performance is significantly higher at both pen and section level with the application of the 2/0 time window than with the 1/0 time window. This indicates that an alarm generated within a 2/0 time window is more prone to be true than an alarm generated within a 1/0 time window. Both time windows would likely be of managerial value, though, and the better choice for an implementation of the detection system would depend on the preferences of the manager of the individual herd.

Thus, alarms at pen level would allow the manager to target preventive interventions in specific pens whereby a spreading of the condition throughout the section may be avoided. Alarms at section level may, however, be generated on the basis of either large changes in the drinking pattern of one pen, or by simultaneous changes in multiple pens within the section. If a section-specific alarm

Table 4.3: AUC (area under curve) for the HHH model version (all three harmonic waves defined at herd level) for both Herd A (commercial finishers) and Herd B (research centre weaners) at three spatial levels given three lengths of time windows. Herd level = any pen in the herd, Section level = a specific section in the herd, and Pen level = a specific pen in the herd. 3/0 time window covers tree days before the event and zero days after the event, 2/0 time window covers two days before the event and zero days after the event, 1/0 time window covers one day before the event and zero days after the event.

Area-specific	Herd A			Herd B		
level	3/0	2/0	1/0	3/0	2/0	1/0
Herd	0.9358	0.9194	0.8013	0.9842	0.9734	0.8878
Section	0.8882	0.8708	0.8144	0.8715	0.8576	0.7705
Pen	0.8878	0.8701	0.8164	0.7671	0.7348	0.6871

is caused by multiple simultaneous pen-specific alarms, then one section-specific alarm would be communicated instead of multiple pen-specific alarms.

4.3.3 *Findings Herd B*

In Herd B, no single model version provided the highest AUC across all spatial levels (results in Paper III). However, the HHH model version obtained the highest detection performance at herd and section levels given all time window settings, which coincides with the findings in Herd A. The highest detection performance at pen level was obtained by the SPP model version when the 3/0 and 2/0 time windows were applied, and by the PPP model version when the 1/0 time window was applied.

Both the SPP model version and the PPP model version define almost individual drinking patterns in each pen of the herd (see Table 3.1), and these results may be influenced by the night activity in the pens, which were discussed in Section 4.2.4. The performances are generally poor for all model versions detecting events at pen level in Herd B, regardless of the length of the time window, as shown for the HHH model version in Table 4.3. Pen level settings should therefore be tested on data from another weaner herd in order to clarify whether the results are specific for Herd B, having only 15 pigs per pen, or if they apply to herds with more pigs per pen as well.

The finding of the HHH model version obtaining the highest detection performance in general in Herd B, coincides with the findings for Herd A. The reasonings for the results coincide as well, and will not be repeated here.

Based on detection performance and managerial value, only one combination of model settings showed satisfying abilities for detecting unwanted events in Herd B. This combination is the HHH model version generating section level alarms within the 2/0 time window (AUC = 0.8576). As seen in Table 4.3, the performance of this combination is a little lower than for the same settings in Herd A (AUC = 0.8708). They are, however fairly high, and both section level alarms and a 2/0 time window constitute settings with high managerial value.

As discussed in Section 4.1.5, performance is measured by comparisons of alarms to a gold standard, which seldom reflect the true state of the animals. In the presented study, water consumption is monitored as an indirect indicator of health and welfare in growing pigs, and the results found in Paper III clearly show that the water consumption contains a high degree of information on the monitored animals. The information is, however, more likely to reflect the general health and wellbeing of the animals, rather than to detect a specific condition.

On that consideration, the interpretation of the area-specific alarms should be that the alarms point out specific focus areas in the herd, not that they predict specific events. Area-specific alarms allow the manager to include knowledge of the pigs in that area, including high risk periods as mentioned in Section 1.2. Hereby a more detailed and thorough focus in the pointed area can be applied, both regarding the pigs, and factors affecting the pigs.

4.3.4 *Alarm handling strategies*

The managerial value of an alarm from a livestock detection system will always be evaluated as a trade-off between the added information on the animals and the extra time spend on attending the alarm. An acceptance of lower detection performances, than the minimum demands defined by Hogeveen et al. (2010), has been suggested in Section 4.1.4. Lower model performances will, however, lead to more alarms being generated, and they may offer a varying degree of information on the animals. In order to prioritize the information, or reduce the generated alarms, an alarm handling strategy can be applied.

Alarms are generated by a two-sided tabular Cusum control chart in Paper III. The output of the dynamic linear model is monitored by the Cusum chart, which generate alarms for any specific pen, any specific section, or for the herd in general. This spatial monitoring allows for an alarm prioritizing and an alarm reducing strategy, which are both founded in simultaneous monitoring of forecast errors in pens and sections. The difference between a prioritization and a reduction of alarms is, that the former ranks certain alarms as more important than other, whereas the latter merely reduces the number of alarms communicated to the manager. In the following, both strategies will be presented.

Alarm prioritizing strategy

The alarm prioritizing strategy is based on the occurrence of alarms in a pen and the corresponding section at the exact same time, t, as illustrated in Figure 4.1. Section level alarms are either caused by a very large deviation in the drinking pattern of a single pen within the section, or by several relatively smaller simultaneous alarms in more pens within the section. Very large deviations in a single pen may be caused by a sudden malfunction in a drinking bowl or a drinking nipple, but deviations may also be caused by a very abrupt changes in the drinking activity in the pen. A number of relatively smaller simultaneous alarms, on the other hand, indicate that the health or wellbeing of a larger number of pigs in a section has changed at the same time. A pen-specific alarm, which has instant impact at section level, is likely to contain more information on the animals than a pen-

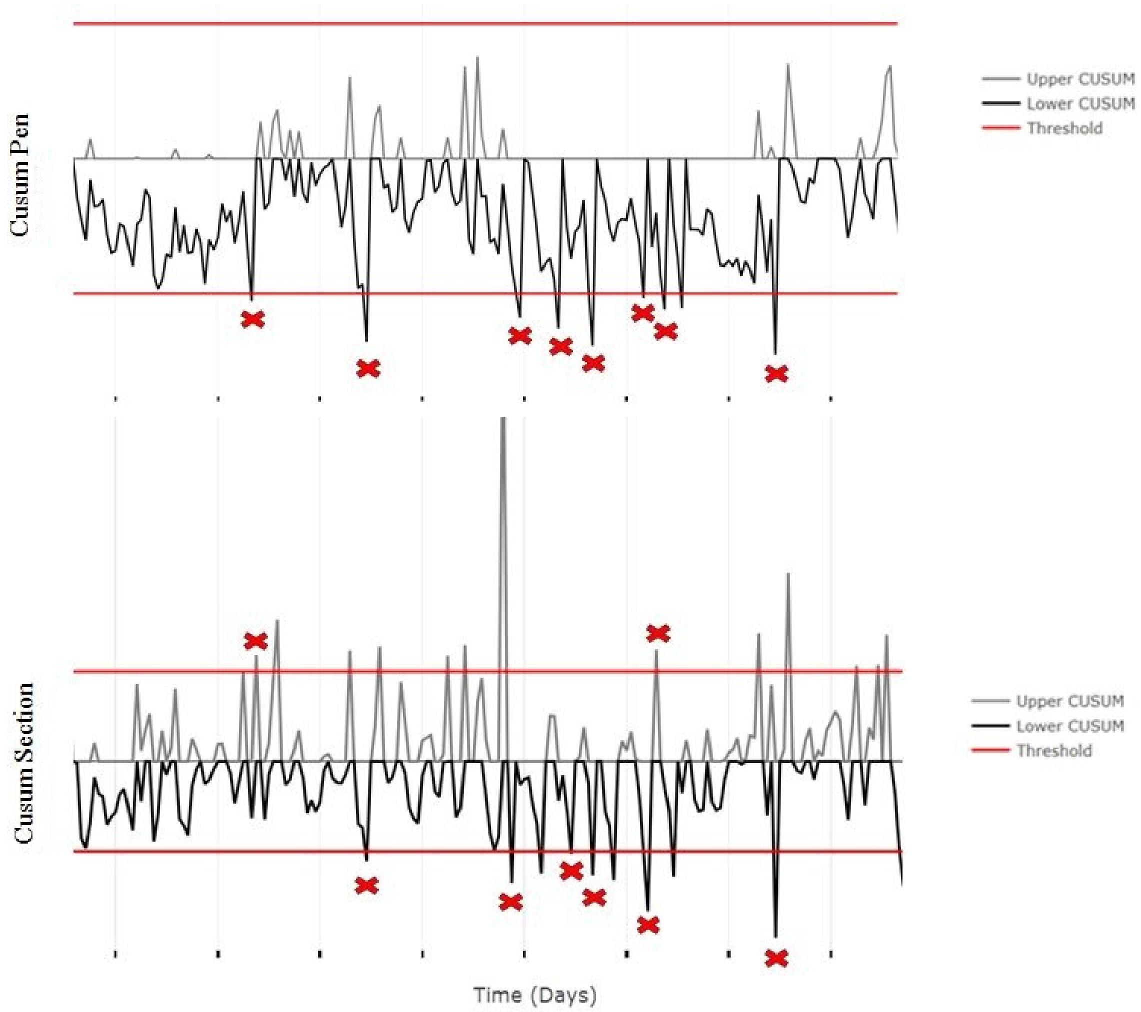


Figure 4.1: Example of a Cusum from one pen (top) and the corresponding section (bottom) with simultaneous alarms. The two red lines mark the upper and lower thresholds. If the threshold is reached or exceeded by the Cusum, an alarm is generated. Alarms marked with an X occur at the exact same hour in the pen as in the section (Figure from Paper III).

specific alarm, which only occur at pen level. Therefore the alarm prioritizing strategy implies that alarms, which occur at the same time at pen level and section level, should be prioritized in the daily management.

Alarm reducing strategy

The alarm reducing strategy is based on the merging of more simultaneous pen-specific alarms into one alarm communicated for the section. Alarms occurring on the same day in multiple pens within the same section, can be merged and communicated as one alarm for the section instead of multiple individual pen-specific alarms. Fewer alarms pointing towards a section may constitute a higher managerial value than a higher number of pen specific alarms. They may, however, also devaluate pen-specific information for the sake of communicating fewer alarm. The preferences in this trade-off will always depend on the animal health status, managerial resources and preferences in the individual herd.

The section alarms generated with this alarm reducing strategy differs from those generated by a section-specific Cusum chart. The alarm reducing strategy merges alarms generated separately on pen-specific vectors of forecast errors, whereas the section-specific Cusum generates alarms based on added forecast error vectors from all pens within the section, as described in Section 3.4.5.

Section-specific alarms may provide sufficient information

In Paper IV, the same Cusum settings were applied to one week's data from all pens within a section in Herd A and in Herd B. Subsequently the number of alarms per pen were counted, and the sum of pen-specific alarms was compared to the number of section-specific alarms for the associated section. For Herd A, two pen-specific Cusum charts generated a total of 6 individual alarms for the week and the section-specific Cusum yielded 4 section alarms. In Herd B, one pen was empty, and therefore three pen-specific Cusum charts generated a total of 42 pen alarms as compared to 8 section alarms generated by the section-specific Cusum chart (results in Paper IV).

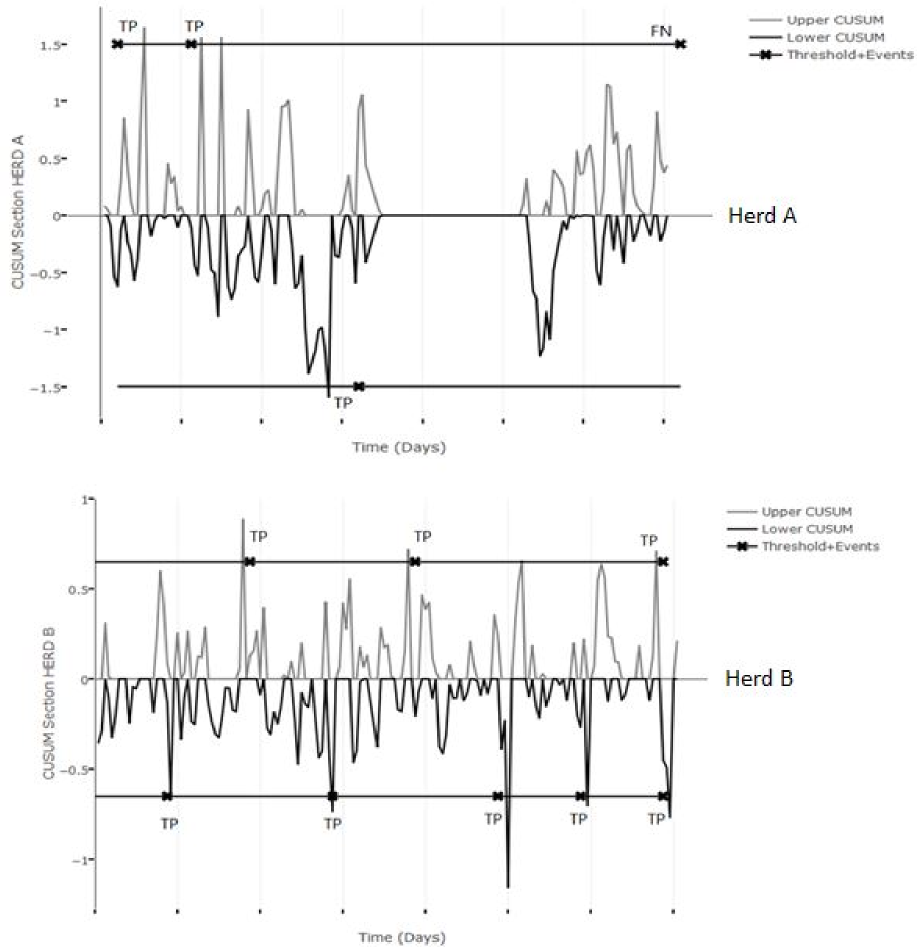


Figure 4.2: Example of a Cusum for one week in a section of Herd A and a section of Herd B. The two horizontal lines mark the thresholds for the upper CUSUM (grey line) or the lower CUSUM (black line). Four events (marked by x on the threshold lines) are registered in Herd A and eight in Herd B. The tabular CUSUM detects three events in Herd A, and eight in Herd B. TP = True Positive, FN = False Negative. The gap around day 5 in the plot is caused by sensor outage (Figure from Paper IV).

Figure 4.2 shows section-specific events and section-specific Cusum charts for the tested week for Herd A and for Herd B. The Cusum charts illustrate remarkably coinciding alarms and events in both herds, which indicate, that section-specific alarms may contain sufficient information on the state of the pigs they represent. This can imply, that the extra information contained in pen-specific alarms is of less relevance than the information in section-specific alarms. The high number of pen-specific alarms in Herd B may reflect the night activity cause alarms to be generated. Since the night activity occurs at different times in each pen, the forecast errors from the pens may even out each other as the pen-specific vectors of forecast errors are added into one vector of section-specific forecast errors.

Although the findings are made on data from the study, they only explore data from one week in one section of each herd. Furthermore no time windows were applied in the example, which would have reduced the number of pen-specific alarms for the comparison.

4.3.5 *Conclusion Papers III and IV*

In Paper III, each of the seven model versions for Herd A and for Herd B are evaluated for their ability to detect outbreaks of either diarrhea or fouling (unwanted events) based on changes in pigs' drinking patterns, at three spatial levels. The findings show that it is possible to generate such area-specific alarms with high predictive accuracies and satisfyingly short time windows. Hereby the findings confirm Hypothesis II.

The HHH model version was found to be the overall better performing model version, which may be due to a higher rigidity in the model parameters, which allow it to register changes in the observed pattern instead of adjusting to them. The 2/0 time window was found to be the better of three in both herds, whereas the shorter, 1/0 time window resulted in performances equal to the 2/0 time window in Herd A, but not in Herd B.

The generating of area-specific alarms offers multiple strategies for handling and communication of alarms to the manager. An alarm prioritizing and an alarm reducing strategy were suggested, though further research is needed to develop and evaluate the managerial value of the two.

CHAPTER 4 REFERENCES

Andersen, H. M. L., E. Jørgensen, and L. J. Pedersen (2016). „Using Evolutionary Operation technique to evaluate different management initiatives at herd level.“ In: *Livestock Science* 187, pp. 109–113.

Aparna, U., L. J. Pedersen, and E. Jorgensen (2014). „Hidden phase-type Markov model for the prediction of onset of farrowing for loose-housed sows.“ In: *Computers and Electronics in Agriculture* 108, pp. 135–147.

Berckmans, D. (2014). „Precision livestock farming technologies for welfare management in intensive livestock systems. (Special Issue: Animal welfare: focusing on the future.)“ In: *Revue Scientifique et Technique - Office International des Epizooties* 33.1, pp. 189–196.

Cavero, D. et al. (2006). „Mastitis detection in dairy cows by application of fuzzy logic.“ In: *Livestock Science* 105.1/3, pp. 207–213.

Chagunda, M. G. G. et al. (2006a). „A model for detection of individual cow mastitis based on an indicator measured in milk.“ In: *Journal of dairy science* 89.8, pp. 2980–2998.

Chagunda, M. G. G. et al. (2006b). „L-lactate dehydrogenase and N-acetyl- beta -D-glucosaminidase activities in bovine milk as indicators of non-specific mastitis.“ In: *Journal of Dairy Research* 73.4, pp. 431–440.

Cornou, C., J. Vinther, and A. R. Kristensen (2008). „Automatic detection of oestrus and health disorders using data from electronic sow feeders.“ In: *Livestock Science* 118.3, pp. 262–271.

Freson, L. et al. (1998). „Validation of an infra-red sensor for oestrus detection of individually housed sows.“ In: *Computers and Electronics in Agriculture* 20.1, pp. 21–29.

Friggens, N. C. et al. (2007). „Estimating degree of mastitis from time-series measurements in milk: a test of a model based on lactate dehydrogenase measurements.“ In: *Journal of dairy science* 90.12, pp. 5415–5427.

Friggens, N. C., M. C. Codrea, and S. Hojsgaard (2010). „Extracting biologically meaningful features from time-series measurements of individual animals: towards quantitative description of animal status“. In: *7th International Workshop on Modelling Nutrient Digestion and Utilisation in Farm Animals*. Ed. by D. Sauvant et al. Wageningen Academic Publishers. ISBN: 978-90-8686-156-9.

Garcia, E. et al. (2014). „Lameness detection challenges in automated milking systems addressed with partial least squares discriminant analysis.“ In: *Journal of dairy science* 97.12, pp. 7476–7486.

Hawkins, Douglas M. (2004). „The Problem of Overfitting“. In: *Journal of chemical information and computer sciences* 44.1. doi: 10.1021/ci0342472 pmid: pp. 1–12.

Hertem, T. van et al. (2014). „Automatic lameness detection based on consecutive 3D-video recordings.“ In: *Biosystems Engineering* 119, pp. 108–116.

Hogeveen, H. et al. (2010). „Sensors and clinical mastitis - the quest for the perfect alert.“ In: *Sensors* 10.9, pp. 7991–8009.

Jensen, Dan B., Henk Hogeveen, and Albert De Vries (2016). „Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis". In: *Journal of Dairy Science* 99.9, pp. 7344 –7361. ISSN: 0022-0302.

Kamphuis, C. et al. (2008). „Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count." In: *Journal of dairy science* 91.12, pp. 4560–4570.

Kamphuis, C. et al. (2010a). „Decision-tree induction to detect clinical mastitis with automatic milking." In: *Computers and Electronics in Agriculture* 70.1, pp. 60–68.

Kamphuis, C. et al. (2010b). „Detection of clinical mastitis with sensor data from automatic milking systems is improved by using decision-tree induction." In: *Journal of dairy science* 93.8, pp. 3616–3627.

Kamphuis, C., H. Mollenhorst, and H. Hogeveen (2011). „Sensor measurements revealed: predicting the Gram-status of clinical mastitis causal pathogens." In: *Computers and Electronics in Agriculture* 77.1, pp. 86–94.

Klir, George J. and Tina A. Folger (1988). *Fuzzy Sets, Uncertainty, and Information*. Prentice Hall.

Maatje, K., R. M. de Mol, and W. Rossing (1997). „Cow status monitoring (health and oestrus) using detection sensors". In: *Computers and Electronics in Agriculture* 16.3, pp. 245–254.

Madsen, T. N., S. Andersen, and A. R. Kristensen (2005). „Modelling the drinking patterns of young pigs using a state space model." In: *Computers and Electronics in Agriculture* 48.1, pp. 39–62.

Maertens, W. et al. (2011). „Development of a real time cow gait tracking and analysing tool to assess lameness using a pressure sensitive walkway: the GAITWISE system." In: *Biosystems Engineering* 110.1, pp. 29–39.

Mein, G. A. and M. D. Rasmussen (2008). „Performance evaluation of systems for automated monitoring of udder health: would the real gold standard please stand up?" In: *Mastitis control: from science to practice. Proceedings of International Conference*. Ed. by T. J. G. M. Lam. The Hague, Netherlands: Mastitis control: from science to practice. Proceedings of International Conference, The Hague, Netherlands; Wageningen Academic Publishers. ISBN: 978-90-8686-085-2.

Mol, R. M. de and W. Ouweltjes (2001). „Detection model for mastitis in cows milked in an automatic milking system." In: *Preventive veterinary medicine* 49.1/2, pp. 71–82.

Mol, R. M. de and W. E. Woldt (2001). „Application of fuzzy logic in automated cow status monitoring." In: *Journal of dairy science.* 84.

Mol, R. M. de et al. (2013). „Applicability of day-to-day variation in behavior for the automated detection of lameness in dairy cows." In: *Journal of dairy science* 96.6, pp. 3703–3712.

Quimby, W. F. et al. (2001). „Application of feeding behaviour to predict morbidity of newly received calves in a commercial feedlot." In: *Canadian Journal of Animal Science* 81.3, pp. 315–320.

Rasmussen, M. D. (2002). „Defining acceptable milk quality at time of milking". In:

Rasmussen, M. D. (2005). „Visual scoring of clots in foremilk." In: *Journal of Dairy Research* 72.4, pp. 406–414.

Somers, J.G.C.J. et al. (2003). „Prevalence of Claw Disorders in Dutch Dairy Cows Exposed to Several Floor Systems". In: *Jounal of Dairy Science* 86, pp. 2082–2093.

Sorensen, C. G. et al. (2010). „Conceptual model of a future farm management information system.“ In: *Computers and Electronics in Agriculture* 72.1, pp. 37–47.

Steeneveld, W. et al. (2010a). „Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems.“ In: *Journal of dairy science* 93.6, pp. 2559–2568.

Steeneveld, W. et al. (2010b). „Simplify the interpretation of alert lists for clinical mastitis in automatic milking systems.“ In: *Computers and Electronics in Agriculture* 71.1, pp. 50–56. DOI: http://dx.doi.org/10.1016/j.compag.2009.12.011. URL: http://www.sciencedirect.com/science/journal/01681699.

Tello, A. S. et al. (2011). „Selection of a golden standard for visual-based automatic lameness detection for dairy cows“. In: *Animal hygiene and sustainable livestock production. Proceedings of the XVth International Congress of the International Society for Animal Hygiene*. Ed. by J. Kofer and H. Schobesberger. Wageningen UR Livestock Research, Lelystad, Netherlands. Vienna, Austria: Tribun EU. ISBN: 978-80-263-0008-3.

Torgo, Luís (2017). *Data Mining with R Learning with Case Studies*. Chapman & Hall.

Weber, N. et al. (2015). „Occurrence of diarrhoea and intestinal pathogens in non-medicated nursery pigs.“ In: *Acta Veterinaria Scandinavica* 57.64, pp. 57–64. DOI: DOI10.1186/s13028-015-0156-5.

Witten, I.H. and E. Frank (2005). *Data mining: practical machine learning tools and techniques*. 2nd edition. Morgan Kaufmann, Elsevier.

5

PAPER I

Published in *Computers and Electronics in Agriculture*: Dominiak and Kristensen (2017)[1]

CHAPTER 5 REFERENCES

Dominiak, K. N. and A. R. Kristensen (2017). „Prioritizing alarms from sensor-based detection models in livestock production - a review on model performance and alarm reducing methods.“ In: *Computers and Electronics in Agriculture* 133, pp. 46–67.

[1]This research was funded by the Danish Council for Strategic Research (The PigIT project, Grant number 11-116191). I further wish to thank the anonymous farmers, the Danish Pig Research Centre and the technical staff (particularly Mads Ravn Jensen) at Aarhus University for installing and supervising the sensors and for taking care of the daily observations in the herds.

Computers and Electronics in Agriculture 133 (2017) 46–67

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Review

Prioritizing alarms from sensor-based detection models in livestock production - A review on model performance and alarm reducing methods

K.N. Dominiak[a,b,*], A.R. Kristensen[a]

[a] HERD - Centre for Herd-oriented Education, Research and Development, Department of Large Animal Sciences, University of Copenhagen, Grønnegårdsvej 2, 1870 Frederiksberg C, Denmark

[b] Department of Animal Science, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark

A R T I C L E   I N F O

Article history:
Received 3 November 2015
Received in revised form 8 November 2016
Accepted 10 December 2016

Keywords:
Sensor
Early warning system
Performance
Sensitivity
Specificity
Automatic monitoring
Livestock production

A B S T R A C T

The objective of this review is to present, evaluate and discuss methods for reducing false alarms in sensor-based detection models developed for livestock production as described in the scientific literature. Papers included in this review are all peer-reviewed and present sensor-based detection models developed for modern livestock production with the purpose of optimizing animal health or managerial routines. The papers must present a performance for the model, but no criteria were specified for animal species or the condition sought to be detected. 34 papers published during the last 20 years (1995–2015) are presented in three groups according to their level of prioritization: "Sheer detection models" based on single-standing methods with or without inclusion of non-sensor-based information (19 papers), "Improved detection models" where the performance of the described models are sought to be improved through the combination of different methods (12 papers) and "Prioritizing models" where the models include a method of ranking or prioritizing alerts in order to reduce the number of false alarms (3 papers). Of the three methods that rank or prioritize alerts; Fuzzy Logic, Naive Bayesian Network (NBN) and Hidden phase-type Markov model, the NBN shows the greatest potential for future reduction of alerts from sensor-based detection models in livestock production. The included detection models are evaluated on three criteria; performance, time-window and similarity to determine whether they are suitable for implementation in modern livestock production herds. No model fulfills all three criteria and only three models meet the performance criterion. Reasons for this could be that both sensor technology and methods for developing the detection models have evolved over time. However, model performance is almost exclusively presented by the binary epidemiological terms Sensitivity (Se) and Specificity (Sp). It is suggested that future research focus on alternative approaches for the output of detection models, such as the prior probability or the risk of a condition occurring. Automatic monitoring and early warning systems offer an opportunity to observe certain aspects of animal health, welfare, and productivity more closely than traditionally accomplished through human observation, and the opportunities for improving animal welfare should continue to be a driving force throughout the field of precision livestock farming.

Contents

* Corresponding author at: HERD - Centre for Herd-oriented Education, Research and Development, Department of Large Animal Sciences, University of Copenhagen, Grønnegårdsvej 2, 1870 Frederiksberg C, Denmark.
E-mail address: knd@sund.ku.dk (K.N. Dominiak).

http://dx.doi.org/10.1016/j.compag.2016.12.008
0168-1699/

1. Introduction

Livestock production has moved from extensive production to intensive production over the last few decades (Sorensen et al., 2010). Society's demand for high-quality animal products is continuously increasing while the number of farmers producing the products is decreasing (Kashiha et al., 2013; Berckmans, 2014). The natural consequence of this situation is a centralization of the production units with increasing numbers of animals at each site (Sorensen et al., 2010; Kashiha et al., 2013).

This centralization, together with the 2008 financial crisis, has changed the conditions of the whole managerial situation leaving the farmer with fewer personnel and less time for each of the daily management tasks creating an increasing market for technical solutions. Technology in livestock production includes automatic monitoring and management information systems (MIS), which gathers available information, and decision support systems (DSS), which analyses the available information, in order to detect and control the health and welfare status of the animals at any given time, by providing early warnings of potential problems (Sorensen et al., 2010; Kashiha et al., 2013; Berckmans, 2014).

Giving the right alarm at the right time is a crucial property of an early warning system, and too many false alarms represent a recurring challenge throughout the field of building models for early warning systems. The false alarms are time-consuming and diminish the trust in the system which in livestock production might lead to the consequences of farmer or personnel either ignoring the alarms from time to time or making personal prioritization of the alarms based on experience, time expenditure, gut feeling and work enthusiasm. In such cases, both animal welfare and gross margin are at risk of being compromised and in order to optimize the benefit of an early warning system for the farmer, a prioritization of alarms must be made ensuring communication of only the relevant alarms to the farmer.

Prioritization of alarms can be done at two levels; either by a reduction in the number of false alarms produced by the early warning system, or by a prioritization of alarms. A reduction in the number of alarms can be done through a satisfying level of performance of the early warning system, while a prioritization of alarms seek to rank true positive (TP) and false positive (FP) alarms. Ranking can be done according to severity of the condition in focus, for example lameness, from those that need immediate attention to those that can be attended within a given period of time. The ranking can be made according to different overall motivations such as animal welfare, costs or production efficiency.

The aim of this review is to evaluate methods for prioritizing sensor-based alarms in livestock production in order to reduce the number of false alarms. The evaluation will be done through a presentation of the different methods described in the scientific literature. Then the advantages and disadvantages of the methods, for their realistic implementation in commercial livestock production, are discussed.

The studies included in this review are of such a variety in terms of study-designs, conditions in focus, and definitions of case (a condition, which should be detected by the model) vs non-case (a condition, which should not be detected by the model), that a true comparison of methods and results are not possible. Therefore, this review does not focus on one species, one condition, or on one type of sensor. Instead, it strives to elucidate the general development of sensor-based detection models with a focus on the prioritizing methods. The challenging task of expressing biological variation through statistical methods at an implementable level of accuracy is hereby sought illustrated.

2. Conceptual framework

2.1. Sensor-based detection systems

The idea of a sensor-based detection system is to automatically detect a condition based on observations from one or more sensors installed in the pen or the barn. Examples of conditions include oestrus, parturition, diseases or impaired productivity. In most cases, the outcome is binary in the sense that the condition is either present or not present at a certain time. The following description assumes a binary outcome, although detection systems with categorical outcomes are also reported (e.g. Cornou and Lundbye-Christensen, 2010; Viazzi et al., 2013).

The basic principles behind a detection system can be described as follows: Assume that a sensor system observes the value of a variable x_t at time $t = 1, \ldots, T$. The variable can either be univariate (i.e. a scalar) or multivariate (i.e. a vector). We shall denote as D_t the set of all observations until time t, i.e. $D_t = \{x_1, \ldots, x_t\}$.

The detection system will typically provide some kind of summary statistic $s_t = f(D_t)$ based on the available information until now. The function f can be very simple, for example $f(D_t) = x_t$ (returning the most recent observation) or $f(D_t) = (x_{t-n+1} + \cdots + x_t)/n$ (returning the average value of the n most recent observations). However, the f function can also be derived through more sophisticated advanced methods like Kalman filtering, neural networks or other computer intensive methods.

The detection is (either literally or conceptually) based on the comparison of the summary statistic s_t to a predefined threshold τ. An alert is given if the summary statistic s_t exceeds the threshold. Thus, at time t, we will either have the event $A_t^+(\tau) : s_t > \tau$ "Alert at time t" or the event $A_t^-(\tau) : s_t \leqslant \tau$ "No alert at time t".

As a very simple example of this framework, assume that we wish to detect a certain disease in an animal. The disease is known to cause fever, so a temperature sensor is attached to the animal. The temperature is logged every hour and transmitted to a computer. In this case $x_1, \ldots, x_t$ are simply hourly temperature measurements. A simple summary statistic would be the current temperature implying that $s_t = f(D_t) = x_t$, but also the average over the last few hours might be relevant.

In order to finish the detection system we need to define a threshold, τ. Assuming that the normal temperature of the animal in question is τ_0, it would be natural to choose a higher threshold $\tau = \tau_0 + \delta$ where $\delta > 0$. It is not straight forward to choose the threshold. It is obvious that if δ is small, many alerts will be given. It has the advantage, that most of the disease cases will be found (*true positives*), but on the other hand, we will also have cases where the temperature is above the threshold for other reasons (oestrus, measurement errors or other conditions). In other words, a low threshold will lead to many *false positive* cases. If, on the other hand, a high threshold is chosen, the number of false positive cases will decrease but on the cost of sometimes not detecting true cases (for instance if they are less severe). Thus, we are at risk of having many *false negative* cases.

This illustrates the general problem in detection methods, namely that there is a built-in conflict between few false positive and few false negative cases. Methods for measuring the performance of detection systems are therefore needed. The traditional approach has been to characterize a detection method by two conditional probabilities known as the *sensitivity* and the *specificity*. For given threshold, τ, the sensitivity, se_τ, and the specificity, sp_τ, are defined as follows

$$se_\tau = \mathrm{P}(A_t^+(\tau)|E_t^+) \tag{1}$$

$$sp_\tau = \mathrm{P}(A_t^-(\tau)|E_t^-), \tag{2}$$

where E_t^+ and E_t^- are the true presence and absence, respectively, of the condition we try to detect.

It should be noticed that all performance indicators introduced so far are specific for the chosen threshold. Since, in many cases, the threshold can be chosen so that the sensitivity becomes 1 (or close to one) it will be at the cost of a lower specificity. It is, therefore, necessary always to look at both primary performance indicators simultaneously.

In order to estimate an over-all performance indicator (independently of a threshold), the *Receiver Operating Characteristic* Curve, *roc*, is often used. The curve is defined by the following parametrization:

$$roc = \{(fpr(\tau), se(\tau)) : \tau \in R\}, \tag{3}$$

where $fpr(\tau) = 1 - sp(\tau)$. The over-all performance indicator is the *Area Under Curve*, determined as

$$auc = \int_{\infty}^{-\infty} se(\tau) fpr'(\tau) d\tau. \tag{4}$$

A perfect system will have an $auc = 1$ so, in general, values close to 1 are preferred.

A study by Aparna et al. (2014) has chosen a completely different approach, where the summary statistic is defined as the expected time to next condition. Thus, if the random variable Θ is the time to next condition, then

$$s_t = f(D_t) = \mathrm{E}(\Theta|D_t). \tag{5}$$

Hence, there is no comparison with a chosen threshold. This seems to be a natural approach in cases where the condition *will* eventually happen (e.g. oestrus or parturition) or will happen with high probability.

An overview of the symbols, concepts and definitions is given in Table 1.

2.2. Estimating the performance of detection methods

Even though Eqs. (1) and (2) define the most common performance indicators it is, in most cases, not possible to calculate them analytically. Instead, they must be estimated from data. A necessary condition is that a *gold standard* allowing us to know the true

Table 1
Conceptual framework and performance assessment of sensor based detection systems.

Symbol	Description	Formula/Condition[a]
x_t	Observation at time $t = 1, \ldots, T$	From sensors
D_t	Set of all observations until now	$D_t = \{x_1, \ldots, x_t\}$
s_t	Summary statistic at time t	$s_t = f(D_t)$
τ	Threshold at time t	Decided
E_t^+	Condition (true) at time t	Gold standard
E_t^-	No condition at time t	Gold standard
$A_t^+(\tau)$	Alert at time t with threshold τ	$s_t > \tau$
$A_t^-(\tau)$	No alert at time t with threshold τ	$s_t \leqslant \tau$
$se(\tau)$	True sensitivity with threshold τ	$se(\tau) = \mathrm{P}(A_t^+(\tau)|E_t^+)$
$sp(\tau)$	True specificity with threshold τ	$sp(\tau) = \mathrm{P}(A_t^-(\tau)|E_t^-)$
$er(\tau)$	True error rate with threshold τ	$er(\tau) = \mathrm{P}(E_t^-|A_t^+(\tau))$
$fpr(\tau)$	False positive rate with threshold τ	$fpr(\tau) = 1 - sp(\tau)$
roc	Receiver Operating Curve	$roc = \{(fpr(\tau), se(\tau)) : \tau \in R\}$
auc	Area Under Curve	$auc = \int_{\infty}^{-\infty} se(\tau) frp'(\tau) d\tau$
TP_τ	Number of true positive cases	$\mathrm{TP}_\tau = \sum_t I(A_t^+(\tau) \cap E_t^+)$
FP_τ	Number of false positive cases	$\mathrm{FP}_\tau = \sum_t I(A_t^+(\tau) \cap E_t^-)$
TN_τ	Number of true negative cases	$\mathrm{TN}_\tau = \sum_t I(A_t^-(\tau) \cap E_t^+)$
FP_τ	Number of false negative cases	$\mathrm{FP}_\tau = \sum_t I(A_t^-(\tau) \cap E_t^-)$
Se_τ	Estimated sensitivity	$\mathrm{Se}_\tau = \mathrm{TP}_\tau/(\mathrm{TP}_\tau + \mathrm{FN}_\tau)$
Sp_τ	Estimated specificity	$\mathrm{Sp}_\tau = \mathrm{TN}_\tau/(\mathrm{TN}_\tau + \mathrm{FP}_\tau)$
SR_τ	Estimated success rate	$\mathrm{SR}_\tau = \mathrm{TP}_\tau/(\mathrm{TP}_\tau + \mathrm{FP}_\tau)$
ER_τ	Estimated error rate	$\mathrm{ER}_\tau = 1 - \mathrm{SR}_\tau$
FPR_τ	Estimated false positive rate	$\mathrm{FPR}_\tau = 1 - \mathrm{Sp}_\tau$
FAR_τ	Estimated false alert rate	$\mathrm{FAR}_\tau = \mathrm{FP}_\tau/T$
ROC	Receiver Operating Curve[b]	$\mathrm{ROC} = \{(\mathrm{FPR}_\tau, \mathrm{Se}_\tau) : \tau \in S\}$
AUC	Area Under Curve	Numerical integration

[a] $I()$ is the indicator function.
[b] S is a set of tested values of τ.

state of the system (whether the condition is present or absent) is available. The gold standard is seen as a perfect test enabling us to observe E_t^+ and E_t^- directly. In practise, the gold standard often consists of human observations which is problematic because of the natural subjectivity in these observations, but often it is the only option. In the following description this problem is ignored. It is simply assumed that each time step can be classified as either E_t^+ or E_t^-.

Given a detection system as described in the previous section, a time series of observations $x_1, \ldots, x_T$ and a gold standard, the detection system can be run with a given threshold, τ, for $t = 1, \ldots, T$. This will result in a time series of events drawn from the following four different combinations of detection result ($A_t^+(\tau)$ or $A_t^-(\tau)$) and true state (E_t^+ or E_t^-):

True positive: $A_t^+(\tau) \cap E_t^+$
False positive: $A_t^+(\tau) \cap E_t^-$
True negative: $A_t^-(\tau) \cap E_t^-$
False negative: $A_t^-(\tau) \cap E_t^+$.

The next step in measuring the performance is to count the number of occurrences of each of the four event combinations. Denoting the numbers as TP_τ, FP_τ, TN_τ and FN_τ, respectively, the *estimated* sensitivity, Se_τ, and specificity, Sp_τ, are calculated as

$$\text{Se}_\tau = \frac{\text{TP}_\tau}{\text{TP}_\tau + \text{FN}_\tau} \tag{6}$$

$$\text{Sp}_\tau = \frac{\text{TN}_\tau}{\text{TN}_\tau + \text{FP}_\tau}. \tag{7}$$

Other similar performance indicators like *success rate* (SR), *error rate* (ER), *false positive rate* (FPR) and *false alert rate* (FAR) are also occasionally estimated (see Table 1 for an overview).

The estimated ROC curve is constructed by choosing a large set $S = \{\tau_1, \tau_2, \ldots, \tau_N\}$ of possible threshold values (where $\tau_1 < \tau_2 < \cdots < \tau_N$). For each $\tau_i \in S$, the sensitivity, SE_{τ_i}, and false positive rate FPR_{τ_i} are estimated and plotted as $(\text{FPR}_{\tau_i}, \text{SE}_{\tau_i})$ for $i \in \{1, 2, \ldots, N\}$. Finally, the AUC is determined by numerical integration.

2.3. The curse of false positives

In traditional diagnostic tests, focus is often on the level of sensitivity, because the test usually is carried out only once. With only one test result available it is therefore very important that as many true disease cases as possible are detected. In sensor based detection systems, on the contrary, tests are carried out continuously (or at least regularly). Accordingly, there will be many opportunities to detect a condition so the demands on the sensitivity can be relaxed. Therefore, the true vulnerable point of sensor-based detection systems is the specificity.

Monitoring sensor data from several different data sources has a built-in risk of generating too many false alarms. This can also be the case when only one time series is monitored. The number of false positives may be a problem, even in systems where the specificity of the detection method is very high. This was for instance a problem with an automatic heat detection method (Ostersen et al., 2010) for sows returning to oestrus that had a specificity around 99%. Nevertheless, the error rate (as defined in Table 1 the ratio of false positive out of all alarms) exceeded 95%. This is a natural consequence of sows returning to oestrus being a relatively rare condition.

The phenomenon is easily illustrated using the notation of Table 1. Assume that the condition being detected occurs with probability p at an arbitrary time t (i.e. the *prevalence* is p). Thus, $\text{P}(E_t^+) = p$. The error rate is the conditional probability $\text{P}(E_t^- | A_t^+(\tau))$. According to Bayes' Theorem, we have

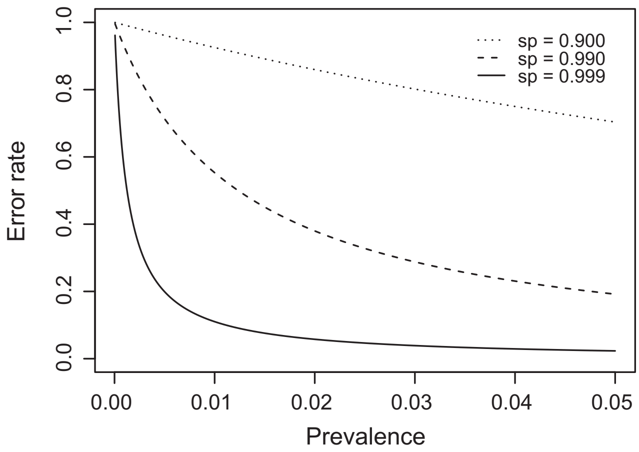


Fig. 1. Error rates as a function of prevalence for three levels of specificity. In all cases the sensitivity is 0.8.

$$\begin{aligned} er(\tau) &= \text{P}(E_t^- | A_t^+(\tau)) \\ &= \frac{\text{P}(A_t^+(\tau)|E_t^-)\text{P}(E_t^-)}{\text{P}(A_t^+(\tau)|E_t^-)\text{P}(E_t^-) + \text{P}(A_t^+(\tau)|E_t^+)\text{P}(E_t^+)} \\ &= \frac{(1 - sp(\tau))(1 - p)}{(1 - sp(\tau))(1 - p) + se(\tau)p}. \end{aligned} \tag{8}$$

Fig. 1 illustrates the error rate as a function of prevalence for three values of specificity and with a sensitivity of 0.8. As it is seen in the figure, even a specificity of 0.99 and a prevalence of 0.01 leads to an error rate above 0.5. In other words, more than half of the alarms will be false positive. If only a specificity of 0.9 is assumed, the error rate will be more than 0.9 with a prevalence of 0.01. When time series from different data sources are monitored simultaneously there is an even higher risk of false alarms.

Only some of the raw alarms will, therefore, require intervention, and it is therefore important to have methods for prioritizing alarms in order to reduce the number of false alarms.

3. Criteria for implementation

According to Hogeveen et al. (2010) three criteria must be fulfilled for a detection model to be implemented in commercial livestock production. These are (A) a high performance in terms of sensitivity (Se) and specificity (Sp), (B) a time window corresponding to the necessary response time for the specific condition, and (C) a high degree of similarity between the study design and the real everyday conditions in commercial farms. The level of value created by the warning system, relative to the investments needed by the farmer for sensors or equipment, could be added as fourth criteria - but first and foremost models fulfilling the three basic criteria must be developed. Throughout this review, the performance criteria is generally given the highest influence when considering the implementability of a model. If the performance is too poor, neither time window nor similarity will be considered further. Should the performance level fulfill the minimum demands (as described in Sections 3.1 and 3.3), the lengths of time windows and the criteria of similarity will be considered according to relevance in the given article.

3.1. Performance considerations

The nature of the condition to be detected must be taken into consideration when defining the level of satisfying performance.

So must the costs and consequences of false alarms in monetary, welfare and production efficiency terms. The performance needed for detecting conditions like oestrus or clinical mastitis (CM), which both need immediate response, is fairly high (Rasmussen, 2002; Ostersen et al., 2010) whereas the demands for detecting conditions like lameness, or impaired daily gain, are considered to be lower, hence reflecting a less urgent condition in some aspects (van Hertem et al., 2013). It is discussed by Pastell and Kujala (2007) if an early warning system, which detects lameness, is rather meant to alert the farmer towards animals that need more focus, than towards animals needing immediate treatment. It is hereby implied that a few false positive alarms have smaller consequences in the detection of lameness than in the detection of for instance CM.

As mentioned in Section 2.3, the prevalence of the given condition highly influences the requirements to the performance. A high Se is desirable when identifying a condition with high prevalence, while a high Sp is necessary when a condition with low prevalence - like CM or oestrus - is sought to be detected (Rasmussen, 2002).

Although the epidemiological terms of Se and Sp are traditionally used for expressing the performance of a detection model, Friggens et al. (2007) state that Se and Sp are of limited value when it comes to monitoring continuous conditions, time series, and progressive scales of conditions. These limitations and the risk-based alternatives to Se and Sp will be discussed in Section 5.3.

Some authors (Firk et al., 2002; Sherlock et al., 2008; Claycomb et al., 2009) have preferred to describe the performance of detection models by SR, ER (Firk et al., 2002), FAR (Sherlock et al., 2008) or FPR (Viazzi et al., 2013). SR is defined as the proportion of true alarms out of all alarms (cf. Table 1) and provides as such an easily interpretable expression of how often the model is right when giving an alarm. Likewise, the ER is the proportion of false positive out of all given alarms. Thus, both SR and ER relate to the number of given alarms, but do not give any information on whether the detection model identifies all cases or has an acceptable level of false negative observations.

FAR, on the other hand, is defined as the proportion of false positive out of the *total number of observations*. This indicator is used by Hogeveen et al. (2010) and Viazzi et al. (2013). Sherlock et al. (2008) suggest that FAR is expressed as the proportion of false positive out of a given, predefined number of observations - for instance 1000 milkings when detecting CM. Communicating to the end user of the alarm system, how many times out of 1000 milkings one must expect a false positive alarm, is easily done, and this interpretation is used by Kamphuis et al. (2008b) and Claycomb et al. (2009).

3.2. The missing gold standard

Throughout the literature, the definitions of case vs non-case are individually set for each study dependent on the study design. In defining a case of CM such different definitions as "presence of clinical signs like clots in the milk or swollen quarters" (de Mol et al., 1997; de Mol et al., 1999), "Somatic Cell Count (SCC) above 100,000 cells/ml or treatment performed" (Cavero et al., 2007) and "one or more alerts given in a defined period around the recorded date of an observed case" (de Mol and Ouweltjes, 2001) illustrate that there is no reference to a generally accepted definition for automated detection of CM, since none currently exists. Mein and Rasmussen (2008) suggest a less stringent definition of a TP case (CM detection) than the one defined in the International Standard (ISO 20966, 2007 in Mein and Rasmussen, 2008, Annex C). This is done as an attempt to agree on a general definition that both maintains the robustness of the gold standard, is practically assessable, and is strengthening the statistics for calculation of the performance of a detection model. The suggestion has not led to a consensus on the matter.

Visual scoring the degree of lameness on a lameness score scale (LS) is widely used as a detection tool. These scorings are often considered the gold standard, although it is a highly subjective method, where the reliability of the scoring result is positively correlated with the experience of the observer (Tello et al., 2011). More than 20 different types of lameness score scales, both discrete and continuous, exist (Tello et al., 2011). Often scorings on a four- or five-point scale are reduced to a three-point trait (Garcia et al., 2014) or even to a binary (Alsaaod et al., 2012), which illustrates the difficulties of ranking lameness in detailed degrees using this method.

Since the terms FP, FN, TP and TN are based on the ability of the detection models to recognize a case or a non-case, it can be argued that with no consensus in the case/non-case definitions, a direct comparison of performance measures is like comparing apples to oranges. This review, however, illustrates the difficulties in obtaining implementable results regardless of the choices of species, conditions and underlying definitions. In Tables 3–5, all inputs are reported with the same terminology as is used in the respective publication when listing methods, variables and performances.

3.3. Performance - minimum requirements

For detection of CM, two minimum requirements for sensitivity are defined in the literature whereas there is only one defined minimum requirement for specificity. In the International Standard (ISO 20966, 2007 in Mein and Rasmussen, 2008, Annex C) a target value for sensitivity is suggested to be 70%, and the target specificity to be above 99%, before a cow is registered on a mastitis attention list. Rasmussen (2002), on the other hand, defines the minimum requirements for sensitivity as 80% and 99% for the specificity, as it is done in Annex C. Since the main reason for building sensor-based detection models is to provide a foundation for better decision support than what human experts can give (Quimby et al., 2001; Kristensen et al., 2010), and since the highest obtained accuracy by human observation is found to be 80% (Rasmussen, 2005), the higher of the two minimum demands to sensitivity is well supported. There is no consensus in the choice of minimum requirement to a threshold though, and both the definitions by Rasmussen (2002) (Hogeveen et al., 2010; Kamphuis et al., 2010b; Huybrechts et al., 2014) and those of Annex C (Kramer et al., 2009; Steeneveld et al., 2010a; Miekley et al., 2012) are used in publications on CM detection.

No standard requirements for performance in lameness detection - or detection of the onset of farrowing - are found in the literature. The performance requirements defined for CM detection will therefore be generally applied when discussing these models.

Some studies define diseases in disease blocks defined as uninterrupted sequence of "days in disease" in association with a detection of the condition (Miekley et al., 2013a). The performance is then expressed in block specificity and block sensitivity (Kramer et al., 2009; Cavero et al., 2007; Miekley et al., 2012; Miekley et al., 2013a). Disease blocks can be defined similar to time win-

Table 2
Distribution of conditions covered by presented detection models. Detection of CM and lameness have had the highest focus overall but also detection of oestrus is well covered. Other diseases, parturition, activity types and weight estimation are all sparsely covered. Some papers cover multiple conditions.

Condition	Animal category	Number of papers
Clinical Mastitis (CM)	Cow	17
Lameness	Cow and sow	14
Oestrus	Cow and sow	9
Other diseases	Cow and sow	5
Parturition	Sow	2
Activity types	Sow	1
Weight estimation	Weaned pigs	1

Table 3
Sheer detection models.

Paper	Focus (E_t^+)	Method ($f(D_t)$)	Sensor	Variables (x_t)	Performance	Performance method	Misc
				Dairy cows			
de Mol et al. (1997)	CM[a] Oestrus Other diseases	Time series Kalman filter	AMS[b]–like Activity tags Feeding troughs with sensors	Milk yield Milk temperature Activity EC[c] Left over concentrates Cow status (for oestrus)	HSe[d] 100 (Sp[e] 84.8) HSp[f] 99.9 (Se[g] 37)	HSe from combination of variables in one of two included farms HSe for other diseases HSp for CM	Milking in parlor but sensors similar to AMS-sensors Different techniques used for Se and Sp Time wondow (CM) 17 days
Maatje et al. (1997)	CM Oestrus Other diseases	Time series Kalman filter	In-line QMC[h] In-line temperature Pedometer	QMC Milk yield Milk temperature Activity level General cow information	HSe 94.5 (Sp 93.5) HSp 98.2 (Se 90)	Multivariate distribution of parameters Cow level	Data from two experimental farms Model based on time series from de Mol et al. (1997)
de Mol et al. (1999)	CM Oestrus Other diseases	Time series Kalman filter	AMS-like sensors Step counter Feeding troughs with sensors Cow status (oestrus)	Milk yield Temperature EC Activity Concentrate intake and ration	HSe 99.6 (Sp 86) HSp 99.4 (Se 65)	HSe for Oestrus HSp for CM	Model developed in de Mol et al. (1997) First cow specific model Two daily milkings
de Mol and Ouweltjes (2001)	CM	Time Series	AMS	EC Milk yield	HSe 100 (Sp -) HSp 99.3 (Se -)	HSe Time-series. Indeterminable EC excluded HSp exponential smoothing. Model default by AMS	Comparison of three models Inconsistency in definitions of true/false cases and in inclusion of 'indeterminable' EC in test. Different techniques for calculating Se and Sp
de Mol et al. (2001)	CM Oestrus Other diseases	Time series Kalman filter	AMS-like sensors but in milking parlor	Milk yield Temperature EC Activity Concentrate intake	HSe 93 (Sp 94) HSp 99.6 (Se 8)	HSe oestrus HSp CM	Comparison of two models and experts' prediction tested in four commercial farms (field condition)
Cavero et al. (2006)	CM	Fuzzy logic	AMS	Milk yield Milk flow rate Time between milkings Milk production rate EC SCC	HSe 92.9 (Sp 93.9) HSp 94.0 (Se 81.1)	HSe broadest CM definition, test data HSp broadest CM definition, training data	Sensitivity set to be minimum 80% Two SCC thresholds SCC weekly from pooled samples Error rates up to 96.1%
Cavero et al. (2007)	CM	Moving Average EWMA[i] Locally weighted regression	AMS	EC SCC[j] Milk yield	HSe 87.9 (Sp 66.6) HSp 81.6 (Se 85.0)	HSe Locally weighted regression HSp EWMA	Three univariate methods based on EC and tested on two SCC-thresholds Block sensitivity
Pastell and Kujala (2007)	Lameness	Probabilistic neural network (PNN)	Balance platform	Weight distribution pr leg	HSe 100 (Sp 57.5)	PNN	5 point lameness score scale reduced to binary Performance evaluated and presented in text and by ROC[k] curve
Kamphuis et al. (2008b)	CM	Fuzzy logic	AMS Inline SCC (ISCC)	ISCC EC FSCC[l]	HSR[m] 32% (FAR[n] 1.2%) Lowest FAR 1.2% (SR[o] 32%)	HSR Multivariate model, largest time window Lowest FAR Multivariate model, largest time window	Alert threshold adjustet to set sensitivity closest possible to 80 Four different time windows used

(continued on next page)

Table 3 (*continued*)

Paper	Focus (E_t^+)	Method ($f(D_t)$)	Sensor	Variables (x_t)	Performance	Performance method	Misc
Kramer et al. (2009)	CM Lameness	Fuzzy logic	Feeding and water troughs AMS neck transponders	Dry matter intake Number of feedings Feeding time Water intake Activity observations	HSe 82.3 (Sp 84.1) HSp 92.1 (Se 75)	HSe (CM) HSp (CM)	Error rates up to 99.5% Sensitivity calculated as block-sensitivity set to be minimum 70%
Maertens et al. (2011)	Lameness	Unspecified. Described as: Specially-developed dedicated kinematic variable-analysis software Linear regression	Pressure sensitive mat	20 kinematic gait variables in space-time Duration of each hoof imprint Location of each hoof imprint	HSe 90 (Sp 100) HSp 100 (Se 90)	Linear regression	Performance reported for gait score 3 Overall 84% of all cows were correctly classified in a three-scale classification
van Hertem et al. (2013)	Lameness	Correlation between binary output and input variables Logistic regression	Accelerometer	Milk yield Neck activity Ruminant time	HSe 89 (Sp 85) HSp 85 (Se 89)	Logistic regression	Milk yield higher correlated than neck activity or ruminant time Only strickly healthy cows included in data set
Kamphuis et al. (2013)	Lameness	Multivariate additive logistic regression	Weight scale Pedometer AMS	Liveweight Activity Milk-order, - yield and - duration	HSe 56.8 (Sp 80) HSp 90 (Se 41)	Multivariate additive logistic regression with AUC minimum 6	5 point lameness score scale reduced to binary Se calculated at fixed Sp levels of 80 and 90
Miekley et al. (2013b)	Mastitis Lameness	Principal Component Analysis (PCA)	Milk meter Pedometer Feeding trough with sensors	Milk yield EC Activity Feed intake Number of feeding visits Feeding time	HSe 87.8 (Sp 61.9) HSp 76.7 (Se 83.3)	HSe for lameness HSp for CM PCA	Disease blocks of day of treatment plus 3–7 days before High error rate (99 for CM, 87.8 for lameness) TP[p] and FP[q] cows/day presented
Viazzi et al. (2013)	Lameness	Decision tree	Video Camera	BMP[r] Back posture	TPR[s] 0.94 (FPR[t] 0.24) FPR 0.04 (TPR 0.25)	TPR (not lame) FPR (lame)	5 point LS[u] scale reduced to 3 points (FPR) or binary scale (TPR)
van Hertem et al. (2014)	Lameness	Ordinal multinomial logistic regression Nomial mulitinomial logistic regression Linear regression	3D-camera Photocell	Back-Posture-Measurement (BPM) Locomotion score (LS)	HSe 54.9 (Sp 90.4) HSp 94.1 (Se 47.1)	HSe Linear regression HSp Ordinal multinomial logistic regression	5-point LS transformed to binary Camera-data collected during night-milkings due to sensor-sensitivity to sunlight
				Sows			
Bressers et al. (1995)	Oestrus	Threshold	Camera	Duration of visits to a boar Frequency of visits to a boar Combination of the two First standing response	95% of sows in oestrus detected in time for service Many false positive	Frequency of visits to a boar	Few details described in paper Oestrus validated by farm checklist

Freson et al. (1998)	Oestrus	Canonical discriminant analysis Logistic regression	Infra-red sensor	Body movement Behavior	Se 79 (Sp 68)	Sensor performance - not model performance	TP and FP distinguished with probability of 99.9% Not possible to discriminate between TN[v] and FN[w]
Cornou and Lundbye-Christensen (2010)	Classify activity types and - categories	Uni- and mulitivariate DLM[x] MPKF[y]	Accelerometer Camera	Feeding Rooting Walking Lying sternally Lying laterally Active or Passive	HSe 96 (Sp 96) HSp 96 (Se 96)	Multivariate DLM with free variance structure Activity category	Only results from one of four model types presented

[a] Clinical Mastitis.
[b] Automatic Milking System.
[c] Electric Conductivity.
[d] Highest sensitivity reported in the article.
[e] Specificity. (Sp) is the specificity corresponding to HSe.
[f] Highest specificity reported in the article.
[g] Sensitivity. (Se) is the sensitivity corresponding to HSp.
[h] Quarter Milk Conductivity.
[i] Exponentially Weighted Moving Average.
[j] Somatic Cell Count.
[k] Receiver Operating Curve.
[l] Laboratory determined SCC.
[m] Highest Success Rate.
[n] False Alert Rate.
[o] Success Rate.
[p] True Positive.
[q] False Positive.
[r] Body Movement Pattern.
[s] True Positive Rate.
[t] False Positive Rate.
[u] Lameness Score.
[v] True Negative.
[w] False Negative.
[x] Dynamic Linear Model.
[y] Multi Process Kalman Filter.

Table 4
Improved detection models.

Paper	Focus (E_t^+)	Improvement	Method ($f(D_t)$)	Sensor	Variables (x_t)	Performance	Performance method	Misc
					Dairy cows			
Liu et al. (2009)	Lameness	Logistic regression	B-spline transformation of limb movement variables (LMV)	Force Load cells	All variables at both limb- and cow level: Peak ground reaction force Average ground reaction force Stance time of a limb Ground reaction force integral	HSe[a] 100 (Sp[b] 100) HSp[c] 100 (Se[d] 100)	HSe 15 knots HSp 15 knots	15 knots is highest degree of freedom LS[e] reduced from 5 point scale to binary No difference in accuracy related to three degrees of freedom Validated by "leave-one-out" 260 times
Kamphuis et al. (2010a)	CM[f]	Decision-tree	Pruning variated Cost matrix	AMS[g]	EC[h] Color Visual scoring	HSe 56.7 (Sp 93.1) HSp 99.8 (Se 5.2)	HSe 5 nodes. Low cost for false classification HSp 26 nodes. High cost for false classification	Both sensor and non-sensor information used to identify positive cows
Kamphuis et al. (2010b)	CM	Decision-tree	J48[i] Bagging Boosting	AMS	EC Color Milk yield Visual scoring	HSe 77.8 (Sp 97.9) HSp 99 (Se 71.4)	J48 combined with bagging	Herd-specific test data Sp fixed at 99% and 97.9%
Miekley et al. (2012)	Mastitis Lameness	Wavelet filtering	CUSUM[j] Selfstarting CUSUM	AMS Pedometer	EC Activity	HSe83.6 (Sp 59.2) HSp 85.5 (Se 63.5)	HSe (CM) CUSUM HSp (lameness) Selfstarting CUSUM	Block-sensitivity at minimum 70% Error rates up to 99.6% Diseases defined as disease blocks
Miekley et al. (2013a)	Mastitis Lameness	Wavelet filtering Vector autoregressive model (VAR)	Multivariate cumultative sum (MCUSUM)	Milk meter Feeding troughs with sensors Pedometer	EC Milk Yield Feeding pattern Activity	HSe 78.9 (Sp 80.4) HSp 81.0 (Se 74.2)	HSe (CM) VAR + MCUSUM HSp (lameness) Wavelet filtering + MCUSUM	Block-sensitivity at minimum 70% Error rates op to 99.6% Diseases defined as disease blocks
de Mol et al. (2013)	Lameness	Quadratic trend models	Models fitted with DLM[k]	Accelerometers AMS Concentrate feeder	10 activity variables Number of milkings Number of refusals Milk yield Concentrates left over	HSe 100 (Sp 98.5) HSp 98.5 (Se 100)	Combined data sets Threshold for lame: 5 4 alerted variables required	Discusses customization through adjustments in the discount factor of the DLM and the threshold for the Bayes factor of the DLM
Huybrechts et al. (2014)	CM	Time series Autoregressive moving average	Synergistic Control Concept CUSUM Shewhart control chart	AMS	EC Milk yield	HSe 63 (Sp -)	CUSUM	Specificity not reported
Garcia et al. (2014)	Lameness	Partial least squares discriminant analysis (PLS-DA)	Logistic regression Backward variable selection of originally 320 variables	AMS Activity tag	Parity 1: 17 variables Parity 2: 28 variables (variables not specified)	HSe 79 (Sp 83) HSp 83 (Se 79)	Model for second parity	70% of observations (gait score 2 and parity 3) excluded Validated by "leave-one-out" 331 times
					Sows/Pigs			
Cornou et al. (2008)	Oestrus Lameness Other health disorders	Univariate DLM	CUSUM V-mask	ESF[l] Eartag	Individual eating rank	HSe 75 (Sp 95.4) HSp 95.4 (HSe 75)	(Oestrus) DLM and V-mask	Study performed in three herds. Highest overall performance reported in this table. FP[m] 2–3 times higher with this method than in ESF alert list Discusses customization through adjustments in the discount factor of the DLM
Ostersen et al. (2010)	Oestrus	Multiprocess DLM DGLM[n] Bayes Theroem	First order Markov properties to the DLM	RFID[o] eartag	Duration of visits to a boar Frequency of visits to a boar Combination of the two Back preassure test	HSe 89.2 (Sp 96.9) HSp 99.4 (Se 55.6)	HSe Combined model (Bayes theorem) HSp Duration model (Multiprocess DLM)	Duration alone exceeds both combined model and frequency alone Error rate for HSe is 97.1% Error rate for HSp is 92% Discusses customization through adjustments in the discount factor of the DLM

Cornou and Lundbye-Christensen (2011)	Parturition	MPKF[p]	DGLM CUSUM	Accelerometer Camera	3 activity parameters 4 passivity parameters Straw/no straw	HSe 100 (Sp 100) HSp 100 (Se 100)	CUSUM	Model over fittet (recognized by authors) Some alarms appearing 0 (zero) hours before farrowing
Kashiha et al. (2014)	Weight estimation	Linear regression Mixed effects (non-linear)	Transfer Function	Cameras Weight scale	Top-view pig area Body weight	Highest R^{2q} 97.5%	Transfer Function	48 Transfer Function models calculated Only one presented in paper

[a] Highest sensitivity reported in the article.
[b] Specificity. (Sp) is the specificity corresponding to HSe.
[c] Highest specificity reported in the article.
[d] Sensitivity. (Se) is the sensitivity corresponding to HSp.
[e] Lameness Score.
[f] Clinical Mastitis.
[g] Automatic Milking System.
[h] Electric Conductivity.
[i] Default AMS algorithm.
[j] Cumulative Sum.
[k] Dynamic Linear Model.
[l] Electronic Sow Feeding.
[m] False Positive.
[n] Dynamic Generalized Linear Model.
[o] Radio Frequency Identification.
[p] Multi Process Kalman Filter.
[q] Coefficient of determination R^2.

dows (Kramer et al., 2009), which will be discussed in the next subsection. A reason for the use of disease blocks can be to focus on early detection (Kramer et al., 2009; Miekley et al., 2013a) but it is important to notice that by calculating block sensitivity instead of sensitivity at day level or even at case level, information on the number of successive alerts is neglected (de Mol et al., 2013). This can cause the sensitivity of the models to be higher since the number of observations registered by the model is reduced, and the chance of TP is increased.

3.4. Time window

Time windows define time frames for conditions in livestock production. The gold standard is the true clinical status, and the time window can be defined as the minimum expected length of the true clinical status (Sherlock et al., 2008). The length of a time window can be based on direct or on indirect indicators. Direct indicators are for example SCC (Hojsgaard and Friggens, 2010), laboratory analyzed hormone levels (de Mol et al., 1997) or visual observations by the farmer for CM in cows (de Mol et al., 1997; Kamphuis et al., 2010b; Miekley et al., 2013a). Indirect indicators can be back pressure test for oestrus in sows (Cornou et al., 2008), changes in feeding behavior for lameness in cows (Quimby et al., 2001) or animal activity as in Cornou et al. (2008) and Cornou and Lundbye-Christensen (2011). Time windows can overlap if the condition is occurring multiple times with short intervals. In such cases, the time windows can be merged into disease blocks (Cavero et al., 2006; Kramer et al., 2009; Miekley et al., 2012; Miekley et al., 2013a), and Se and Sp of disease blocks can be reported as block sensitivity and block specificity (Cavero et al., 2007; Kramer et al., 2009) which might have both advantages and disadvantages as mentioned in the previous subsection.

If a condition is present and an alarm occurs within the time window, the alarm is TP. If an alarm occurs before or after the defined time window, it is considered FP. On the other hand, if a condition is present but no alarm occurs within the defined time window, the situation is FN whereas it is TN if no alarm occurs in the absence of the condition of interest (Fig. 2).

The length of the time window has great influence on the performance of the detection model. A very long time window like 17 days for CM (de Mol et al., 1997) heightens the chance of TP and increases the performance of the alarm system because all alarms occurring within the time window are classified as TP. It can be argued that a system that generates an alarm anywhere between 1 and 10 days before to a week after an intervention is needed, might be of little practical use to the farmer. A short time window, like 6 h for oestrus detection (Ostersen et al., 2010) or during the very milking in CM detection, would generate alarms of great use as a managerial tool (Kamphuis et al., 2010a) for conditions that require fast intervention. But such short time windows increase the demands to the accuracy of the model.

3.5. Similarity between models and real life

Similarity between the study population and commercial livestock production populations is of utmost importance if the developed detection models should have a chance of performing well under field conditions and later be implemented. Three reasons for dissimilarities between the study population and commercial livestock populations are: a narrow data set that does not depict the variety of commercial farms, indistinct definitions of case (TP) and non-case (TN) in the study design, and the capability of the model to handle missing data (Hogeveen et al., 2010). If these criteria are not fulfilled, the risk of a disappointing level of detection performance in a commercial herd is high (Hogeveen et al., 2010).

Table 5
Prioritization models.

Paper	Focus (E_t^+)	Method ($f(D_t)$)	Input	Sensor	Variables (x_t)	Performance	Performance method	Misc
					Dairy cows			
de Mol and Woldt (2001)	Reduction of FP[a] in detecting CM[b] and oestrus	Fuzzy logic. Default and manually optimized	Results from statistical models presented in de Mol and Ouweltjes (2001) (CM) and de Mol et al. (1997) (oestrus)	AMS[c]	Milk yield Temperature EC[d] Activity	HSe[e]100 (Sp[f] 99.75) (CM) HSp[g] 99,75 (Se[h] 100) (CM) HSe 79 (Sp 98.1) (oestrus) HSp 99.4 (Se 66) (oestrus)	CM: Fuzzy Logic FP reduced from 1266 to 64 Oestrus: Fuzzy logic optimized manually and by neurofuzzy technologies FP reduced from 384 to 123	CM: Same data set used for learning and testing High selection in included cows Oestrus: Manual optimization minimal improvement Neurofuzzy no improvement
Steeneveld et al. (2010a)	Identification of alerted cows that need further investigation for CM	Naive Bayesian Network (NBN) Logistic regression	Alert list from AMS Non-AMS cow-information AMS-alert information Combination of non-AMS and AMS information	AMS	EC Color alert CM Color alert Abnormal milk Milk yield Cow specific information	The number of FP alerts reduced by 57 % 10% of TP[i] alerts missing	NBN combining both sensor and non-sensor information	Default performance of AMS in the study does not meet minimum demands of satisfying detection performance
					Sows			
Aparna et al. (2014)	Time to farrowing Probability of farrowing	Hidden phase-type Markov		Water valve Photo-cells Camera	Water consumption Activity Behavior Time of day Time since mating	Estimated time to farrowing SD[j] 4.6 h Probabiliy of farrowing SD 4.5 h Threshold 12 h 97% True warnings	Water consumption and activity in combination gave best result for both time estimation and probability calculation	Condition is known to happen Predictions based on herd specific parameters

[a] False Positive.
[b] Clinical Mastitis.
[c] Automatic Milking System.
[d] Electric Conductivity.
[e] Highest sensitivity reported in the article.
[f] Specificity. (Sp) is the specificity corresponding to HSe.
[g] Highest specificity reported in the article.
[h] Sensitivity. (Se) is the sensitivity corresponding to HSp.
[i] True Positive.
[j] Standard Deviation.

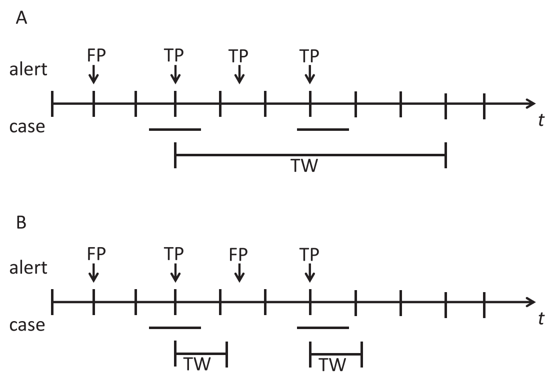


Fig. 2. A long time window (A) can cover more successive alerts hence affecting the performance of the detection model in terms of Se and Sp as well as success rate and error rate. Alerts occurring within the range of the time window are all considered TP. A shorter time window (B) allows for a more detailed classification of all alerts and calculation of model performance. t: time. TW: Time window. TP: True positive. FP: False positive.

A detection model must be validated externally to prove its accuracy under conditions other than the ones it is created under. A high degree of accuracy is reflected in high sensitivity, specificity and reproducibility (Liu et al., 2009).

When evaluating models with promising performances, it is highly relevant to include the validation method to get a more fulfilling picture of the potential for implementation. Financial limitations or different types of deadlines can be reasons for designing the study validation in a way that does not meet the similarity criteria - and it might be of greater importance to build a model first, and then validate it under conditions less challenging than in herds representative for the average production form in the given area of interest.

The strongest validation is on data, which is completely independent from the data set used for training and learning the model, as for instance data coming from another herd. If it is not possible to obtain suitable data from an independent herd, and if the data set is large enough, validation can be done by dividing the original data set into test data, learning data and validation data (Witten and Frank, 2005). Often the data set is too small for such a division, and other methods must be considered. A commonly accepted validation method is a 10-fold cross validation as used by Viazzi et al. (2013). With this method, the data set is randomly divided into ten subgroups, one subgroup is then retained as validation data, and the model is trained on the remaining nine subgroups. The validation is strongest with this method, when the process is repeated ten times, each time with a new subgroup used for validation data, although this is not always done (Witten and Frank, 2005).

Another validation method, which is used by Liu et al. (2009), is "leave-one-out" cross validation. This method is to some extend similar to the 10-fold cross validation, only it is n-fold, where n is the number of observations/animals in the data set. The validation is performed n times, with each observation left out in turn, and the rest of the data set used as training data (Witten and Frank, 2005). Both cross-validation methods mentioned above are relatively narrow in an implementation aspect due to the high degree of dependency between training and validation data.

Basing a model on data from a few animals (Cornou and Lundbye-Christensen, 2010, 2011; Aparna et al., 2014), animals from a single herd (Bressers et al., 1995; Ostersen et al., 2010; Viazzi et al., 2013; van Hertem et al., 2013, 2014; Garcia et al., 2014) or from herds where the managerial status differs from the average commercial herd, as might be the case in a research herd (de Mol et al., 1997, 2013; de Mol and Ouweltjes, 2001; Cavero et al., 2006, 2007; Pastell and Kujala, 2007; Kamphuis et al., 2008b; Kramer et al., 2009; Steeneveld et al., 2010a; Maertens et al., 2011; Miekley et al., 2012, 2013a,b; Kashiha et al., 2014) can have a high impact on the similarity between the study population and commercial herds. This is either because the biological variety of the whole herd is poorly represented in the small study population, or because the routines are adjusted according to demands of the study design. In the case of research herds, the available resources might differ from what is possible in commercial herds.

The similarity of a model is also highly affected in studies where data is collected in herds with extraordinary high/low prevalence of the condition of interest compared to prevalence in average commercial herds (Miekley et al., 2013a; van Hertem et al., 2013, 2014). And the same is true for studies where animals from the same herd are divided into subgroups in order to define learning data and test data (Kramer et al., 2009; Cornou and Lundbye-Christensen, 2010) since this approach ignores any herd-specific correlation, such as genetics or managemental factors.

An obvious reason for not validating the model under field conditions, even though it strengthens the model, is that it can be very time consuming (de Mol et al., 2001; Nielsen et al., 2005). This is often the reason for cross validating on a subgroup of the study population (van Hertem et al., 2014; Viazzi et al., 2013) or using the same data for training and validating the model (Aparna et al., 2014; de Mol and Woldt, 2001; Liu et al., 2009).

The definition of case/non-case can - as previously mentioned - be very individual in some studies (de Mol et al., 1997, 1999; Cavero et al., 2006; Miekley et al., 2012; Kamphuis et al., 2013; Garcia et al., 2014) whereas other studies use definitions and routines that are already used by the personnel in the farm where the data is collected (Maatje et al., 1997; de Mol and Ouweltjes, 2001; Kamphuis et al., 2010a,b; Miekley et al., 2013b; van Hertem et al., 2013, 2014; Huybrechts et al., 2014).

Since it is common in commercial production herds that data are missing at a more or less influential level, a detection model must be able to handle missing data as well. In some studies, data sets with missing data are left out during the model developing process for different reasons (Pastell and Kujala, 2007; Steeneveld et al., 2010a; Maertens et al., 2011; de Mol et al., 2013; van Hertem et al., 2013, 2014; Garcia et al., 2014) whereas other models are based on incomplete - but more realistic - data sets from commercial farms (Bressers et al., 1995; Liu et al., 2009; Kamphuis et al., 2010a,b; Miekley et al., 2013b; Huybrechts et al., 2014), hence showing a higher level of similarity.

4. Criteria for inclusion in this review

4.1. Primary criteria

Papers included in this review are all peer-reviewed and present sensor-based detection models developed for modern livestock production with the purpose of optimizing animal health or managerial routines. Papers on models that are based on parameters analyzed in laboratories (Barkema et al., 1998; Nielsen et al., 2005; Chagunda et al., 2006; Friggens et al., 2007; Steeneveld et al., 2008; Hojsgaard and Friggens, 2010), parameters assessed by humans (Barkema et al., 1998; Steeneveld et al., 2010b), or where the condition in focus is artificially applied to the animal as a part of the study design (Milner et al., 1996; Abell et al., 2014) are therefore not included.

Papers included must furthermore present results from a performance analysis. Papers where methods for detecting, monitoring or assessing parameters for early warning systems are developed, tested or evaluated, but the results are presented as the method having a future potential, are therefore not included. This criterion leaves out several studies (Bressers et al., 1994; Moshou et al., 2001; White et al., 2004; Madsen and Kristensen, 2005; Madsen et al., 2005b; Oliviero et al., 2008; XiangYu et al., 2008; Leroy et al., 2008; Ferrari et al., 2010; Kristensen and Cornou, 2011; Tanida et al., 2011; Hoffmann et al., 2013; Kashiha et al., 2013; Cornou and Kristensen, 2014b; Porto et al., 2014; Abdanan Mehdizadeh et al., 2015; Dutta et al., 2015).

4.2. Conditions detected

A variety of conditions are sought to be detected in papers included in this review. Some papers present models detecting more than one condition (de Mol et al., 1997, 1999) or several methods for detecting the same condition (Cavero et al., 2007; van Hertem et al., 2014). Some papers combine two methods in order to improve the overall performance (Kramer et al., 2009; Kamphuis et al., 2010a; Cornou and Lundbye-Christensen, 2011; Huybrechts et al., 2014), or to prioritize the outcome of the early warning system (de Mol and Woldt, 2001; Steeneveld et al., 2010a; Ostersen et al., 2010). Table 2 shows the distribution of conditions covered. As it appears from the table, detection of CM and of lameness have had the highest focus overall.

4.3. Sensor types

Multiple sensor types are included in this review representing the technological evolution through the last two decades (from 1995 to 2015). Data from automatic milking systems (AMS) form the basis for the vast majority of sensor-based detection models, but a variety of other sensor types are included as well. Sensors for monitoring movement include video cameras, different 2D movement sensors (pedometers and neck transponders), and 3D movement sensors (accelerometers and pressure sensitive sensors like force plates and load cells). Other sensor types (flow meters, feeding troughs with sensors, weight scales and climate computers) also provide valuable information in several studies.

4.4. Methods - presentation

The included papers are presented in three groups according to their level of prioritization:

Group 1 (Table 3): Sheer detection models based on single-standing univariate or multivariate methods with or without the inclusion of non-sensor-based information.
Group 2 (Table 4): Improved detection models where the performance of the described models are sought to be improved through the combination of different methods.
Group 3 (Table 5): Prioritizing models where the model includes a method of ranking or prioritizing alerts in order to reduce the number of false alarms.

In many studies, performance indicators are reported several times due to different thresholds or different subgroups of animals. Therefore an approach has been taken in order to compare the highest level of performance obtained by any method under any circumstances given in the relevant study. In the tables, the notations HSe and HSp are used. HSe is the highest sensitivity achieved in the study, and the specificity in brackets is the corresponding specificity. Equivalently, HSp is the highest specificity achieved in the study, and the corresponding sensitivity is shown in brackets.

The notations "HSe x_m (Sp y)" and "HSp y_m (Se x)" are mathematically defined as

$$(x_m, y) = (\mathrm{Se}_{\tau_x}, \mathrm{Sp}_{\tau_x}), \quad \tau_x = \arg\max_{\tau}\{\mathrm{Se}_\tau | \tau \in S\} \tag{9}$$

$$(x, y_m) = (\mathrm{Se}_{\tau_y}, \mathrm{Sp}_{\tau_y}), \quad \tau_y = \arg\max_{\tau}\{\mathrm{Sp}_\tau | \tau \in S\}, \tag{10}$$

S is the set of thresholds tested in the study.

4.5. Literature search strategy

For the initial search the following keywords were used: automatic monitoring, livestock production, sensors, ranking, prioritizing, alarms and detecting. These keywords were then combined with words like mastitis, lameness, estrus, gain, cow, sow, and broiler. From those basic searches, backward searches were done through references and bibliographies of relevant authors. The databases used for the searches were Ovid (CAB Abstracts, Web of Science, Agricola) and Sciencedirect in the period from November 2014 to June 2015.

5. Method description

In this section, the methods used for building detection models in the reviewed papers are described according to their level of prioritization. In both groups 1 *Sheer detection models* (Table 3) and 2 *Improved detection models* (Table 4) some papers present a technique where the level of one or both performance parameters are fixed (Kamphuis et al., 2010b; Kamphuis et al., 2013), or defined with a minimum level (Cavero et al., 2006; Kramer et al., 2009; Miekley et al., 2012, 2013a) when doing performance analyses. With a fixed parameter, it is possible to calculate the corresponding threshold for detecting a condition under given circumstances, and hereby reduce the number of false alarms generated by the detection model. However relevant, according to the alarm-reducing characteristics, this technique is not a part of the construction of the detection model and will not be described further.

5.1. Sheer detection models

An overview of the sheer detection models (Group 1) identified for this review is shown in Table 3.

5.1.1. Sheer detection models based on AMS sensors
Statistical methods used in models based on data from AMS (or AMS-like) sensors all fall into one of the following four categories: Time series with Kalman filter (de Mol et al., 1997; de Mol et al., 1999; de Mol and Ouweltjes, 2001), local regression, moving averages (Cavero et al., 2007), and fuzzy logic (Cavero et al., 2006; Kamphuis et al., 2008b; Kramer et al., 2009). de Mol et al. (1997, 1999) present a multivariate cow-dependent approach and an AutoRegressive Integrated Moving Average (ARIMA) for analyzing time series with Kalman filter. Later de Mol and Ouweltjes (2001) use an unspecified time series model, where milking intervals and milking frequencies are included as variables. The specificity for CM detection in de Mol et al. (1997) is based on milk sampled with a two month interval, and cows with no CM pathogens or elevated SCC counts in any samples during the study period were defined as TN. This means that a TN cow with one or more alarms was considered FP. This case definition ignores any CM cases which begin and end between two samples, and creates optimal - but unrealistic - conditions for the detection model. The multivariate methods presented by de Mol et al. (1997, 1999) is, however, a novel approach through the incorporation of the animal history and traits, and it is widely implemented in later publications (de Mol and Ouweltjes, 2001; Cavero et al., 2007; Claycomb et al., 2009; Kramer et al., 2009; Kamphuis et al., 2010b; Steeneveld et al., 2010a; Garcia et al., 2014; Huybrechts et al., 2014). General practice at the time of several of these studies was milking in milking parlors, and the use of AMS was in its modest beginning (Kamphuis et al., 2008a; Rutten et al., 2013) which made the inclusion of sensor-based variables limited compared to later studies.

Performances presented in early studies by de Mol et al. (1997, 1999, 2001) and de Mol and Ouweltjes (2001), are fairly high, with either sensitivities or specificities fulfilling the minimum requirements by Rasmussen (2002). The requirements are not met at the same time for both performance measures, though. Not even extremely long time windows, a variety of case definitions, or different techniques for performance analysis in de Mol et al. (1997) led to both parameters meeting the requirements at the same time. Authors agree that the performance of the presented models is too poor for practical implementation and suggest either improvement of both sensors and alert rules (de Mol et al., 1997; de Mol et al., 1999) or addition of temperature sensors that have proven informative in detecting CM (de Mol and Ouweltjes, 2001). Simple control charts and local regression were tested and showed to have poor performance in Cavero et al. (2007), and these methods are only used in combination with other methods in later research (Cornou et al., 2008; Lukas et al., 2009; Cornou and Lundbye-Christensen, 2011; Miekley et al., 2012, 2013a; Huybrechts et al., 2014). Even though a model for milk yield based on time series was suggested already by Deluyker et al. (1990), it was on a general cow level, and de Mol et al. (1999) seem to be the first to model cow-specific "normal" behavior through time series based on sensor data.

A binary classification is bound to misclassify some "grey zone" cows (Cavero et al., 2006). The use of a lower SCC threshold of 100,000 cells/ml in defining healthy/sick cows as used by Miekley et al. (2012); Cavero et al. (2006) raises another concern of employing an artificially high sensitivity (Claycomb et al., 2009) due to too many healthy cows being classified as sick (false positive). Even though the chosen threshold of 100,000 cells/ml is following the definitions from "Deutsche Veterinärmedizinische Gesellschaft e.V." for mastitis, it appears to be too low since a number of papers have reported average bulk tank SCC's from 151,000 cells/ml to >800,000 cells/ml (Maatje et al., 1997; Cavero et al., 2006; Kamphuis et al., 2008b; Claycomb et al., 2009; Kramer et al., 2009; Miekley et al., 2012). Mein and Rasmussen (2008) even suggest that cows could be classified as "true negatives" if the SCC is <200,000 cells/ml and all foremilk samples are without clinical signs.

Fuzzy logic is a method where variables that can obtain multiple lingual values are determined relative to the connection in which they appear. The lingual variables can be "many, few, almost all, several", and they are given a numeric value (degree of membership) between 0 and 1 before they are included in for instance statistical models (Klir and Folger, 1988). When this method is used in models for CM detection (Cavero et al., 2006; Kamphuis et al., 2008b; Kramer et al., 2009), it is applied through three steps of a fuzzy logic system called fuzzification, fuzzy inference and defuzzification:

Fuzzification transforms the sensor-measured input variable to a fuzzy value that is a combination of linguistic interpretation and grade of membership (Kramer et al., 2009).

Fuzzy inference applies a set of IF THEN rules generated on expert knowledge for each trait described by fuzzy values and combines them like IF (all X is Z) AND (no Y is X) THEN (no Y is Z) (Klir and Folger, 1988).

Defuzzification transforms the fuzzy values into one numeric value that is compared with a threshold to determine for instance if a cow has got CM or is healthy (Kamphuis et al., 2008b).

The Fuzzy Logic method was first applied by Cavero et al. (2006) who used it on AMS sensor variables. The thresholds for case definitions were very low, which resulted in high performance (in terms of Se and Sp) and large error rates. Fuzzy logic has been used later for detecting CM with both in-line and on-line SCC (Kamphuis et al., 2008b), and for detecting both CM and lameness (Kramer et al., 2009), but no results suitable for implementation in commercial herds were achieved.

The method is good at representing the form of uncertainty that is naturally imbedded in modeling traits with biological variation. By using the so-called Fuzzy Expert System, crisp values can be fuzzified (Klir and Folger, 1988) before applying rules and defuzzification.

Cavero et al. (2006), Kamphuis et al. (2008b), and Kramer et al. (2009) all use numeric sensor measurements as input variables, and the numeric values are first fuzzyfied to lingual values then defuzzified back to numeric values. This process does not seem intuitively as the most obvious method, but it would be interesting to see Fuzzy logic applied to categorization of lameness degrees in cows since it is a trait with a high degree of biological variation. The Fuzzy logic method is used for combining sensor-based alerts with subjective human judging of CM in the study by de Mol and Woldt (2001), and this will be discussed further in Section 5.3.

5.1.2. Sheer detection models based on behavior and movement sensors

A variety of behavior and movement sensors are used in detecting changes in the behavior or movement pattern of an animal. The changes detected are either due to lameness, or the onset of a condition associated with well known behavioral changes like oestrus or farrowing. Numerous studies employ a variety of techniques for assessing activities. These include pressure platforms measuring weight distribution (Pastell and Kujala, 2007; Oliviero et al., 2008; Pastell and Madsen, 2008; Pastell et al., 2008a,b; Pluym et al., 2013; Mohling et al., 2014), pressure sensitive mats monitoring irregularities in gait patterns (Maertens et al., 2011; Pluk et al., 2012; Van Nuffel et al., 2013), and accelerometers measuring types of activity in two or three dimensions (Cornou and Lundbye-Christensen, 2010; Cornou et al., 2011; van Hertem et al., 2013; Cornou and Kristensen, 2014b). Activity sensors fastened to the animal (Alsaaod et al., 2012; Kamphuis et al., 2013; Miekley et al., 2013b; Dutta et al., 2015) or infrared sensors fastened on inventory (Freson et al., 1998; Aparna et al., 2014) are also used in multiple studies. Although several types of statistical methods have been used for building sheer detection models based on behavior or movement sensors, the performance in general follows the same trend as the sheer detection models based on AMS sensors with either a high sensitivity or a high specificity, and with consensus in the finding that multivariate models outperform univariate (Maatje et al., 1997; Kamphuis et al., 2013; van Hertem et al., 2013).

A study by Miekley et al. (2013b) found missing values are causing up to 30% information loss for some cows when using principal component analysis (PCA), whereas (Pastell and Kujala, 2007) found that other methods, like probabilistic neural network (PNN), handle incomplete data sets better. The use of infrared sensors in detecting onset of oestrus in sows is tested and found inadequate for implementation by Freson et al. (1998) since TN and FN could not be distinguished.

Maertens et al. (2011) present an impressive highest accuracy (HSe 90, HSp 100) in detecting lameness among dairy cows using a spatiotemporal approach. This accuracy is however only on identification of severely lame cows whereas the overall performance of the model is presented as a success rate above 80% without specification of Se, Sp or FP. The spatiotemporal approach is new in lameness detection of livestock animals though, and this is investigated in further research (Pluk et al., 2012; Van Nuffel et al., 2013; Meijer et al., 2014).

Some authors discuss improvements by inclusion of walking speed (Meijer et al., 2014) or longer pressure mats to measure more gait cycles within one measurement (Van Nuffel et al., 2013). The use of sensor mats for lameness detection is still a relatively new area in research, and Pluk et al. (2012) naturally focus more on improving the techniques and choosing the most informative variables and methods instead of on implementation in commercial farms.

The study by Cornou and Lundbye-Christensen (2010) on classifying activity levels of sows prior to farrowing, does not reach the performance defined in Rasmussen (2002), but still the results are remarkable since the corresponding Se and Sp are both 96% as contrary to most other studies that reach either high Se or high Sp. The performance is on identifying a sow in activity (walking, feeding, rooting merged) correctly as opposite to lying down either laterally or sternally. A reliable detection of activity category is valuable in predicting conditions that follow a change in activity level - like oestrus or parturition.

For models built on data from video cameras, infrared cameras or 3D cameras, methods like decision trees or different types of regressions have been used in detecting different conditions. None of the presented models detect with a performance high enough for implementation in commercial farms, and the results by van Hertem et al. (2014) detecting lameness in cows reach neither sensitivities nor specificities matching the definitions in Rasmussen (2002).

Although Viazzi et al. (2013) have simplified the lameness score scale from a 5-point to a 3-point, the performance presented as TruePositiveRate and FalsePositiveRate is too low for implementation. Bressers et al. (1995) only present the success rate and a notice of presence of many false positive in detecting oestrus by monitoring sows' visits to a boar, hereby indicating a high sensitivity and a low specificity. A similar study was later conducted by Ostersen et al. (2010) with more complex methods that will be presented in Table 5.

5.2. Improved detection models

The models in Table 4 all have in common that methods are combined to create an improvement in model performance. The improvements added are different types of control charts (Cornou et al., 2008; Cornou and Lundbye-Christensen, 2011; Miekley et al., 2012; Miekley et al., 2013a; Huybrechts et al., 2014), further development of decision trees (Kamphuis et al., 2010a; Kamphuis et al., 2010b), or of various regression methods (Liu et al., 2009; Kashiha et al., 2014). Combinations of DLM with other methods (Ostersen et al., 2010; de Mol et al., 2013) and partial least squares discriminant analysis fitted by linear regression and improved by reducing the number of variables through backward variable selection (Garcia et al., 2014) are also included.

Combining different types of control charts with wavelet filtering, autoregressive methods, time series, or either univariate or generalized DLMs does not result in a performance high enough for implementation (Cornou et al., 2008; Cornou and Lundbye-Christensen, 2011; Miekley et al., 2012, 2013a; Huybrechts et al., 2014). Using CUSUM in detecting the onset of parturition does however result in both a sensitivity and a specificity of 100% for a subgroup of nine sows based on activity level, and a sensitivity of 100% combined with a specificity of 95% when including all 19 sows in the study (Cornou and Lundbye-Christensen, 2011).

Although the performance obtained by Cornou and Lundbye-Christensen (2011) is impressive, there is an overlap between the individual parameters used in both methods (DGLM and CUSUM) and animal specific reference days which may have increased the performance. This is mentioned by the author as a subject for future change if a large-scale study should be conducted. Few alarms appear at time zero (that is at the actual onset of the farrowing) but the majority of the alarms based on the CUSUM method occur between 12 and 24 h before onset of farrowing (mean 4.7–14.8 h, SD 4.9–9.1 h), while the DGLM method produces alarms in average 15 h before farrowing (SD 4.3–7.5 h). An alarm this long before farrowing with a relatively large standard deviation is suboptimal if the purpose is to be present during farrowing. If the purpose, on the other hand, is to prepare the sow or the farrowing crate in order to reduce piglet mortality as in the later discussed study by Aparna et al. (2014), an alarm long time before would in most cases be sufficient.

Different methods for improving DLMs are presented by Ostersen et al. (2010), de Mol et al. (2013). de Mol et al. (2013) use quadratic trend models fitted with DLM to detect lameness in cows. The presented performance of the model is not suitable for implementation, but the authors mention that both discount factors and threshold for the Bayes factor in the DLM can be adjusted. Adjustments can prioritize a higher or lower Se according to the needs of the end user, which means that the threshold for alarms can be adjusted - or prioritized - according to individual needs.

Ostersen et al. (2010) detect oestrus via both duration of a sow's visit to a boar, the frequency of the visit, and a combination of the two parameters. Ostersen et al. (2010) combine a multiprocess DLM with Markov probabilities of the DLM components in the duration model and develop a DGLM for the frequency model. The detection model combining both duration and frequency is based on Bayes Theorem and calculates a combined probability of the sow being in oestrus.

The multivariate model surprisingly enough performs worse than the univariate duration model. An explanation for this finding could be that the duration model includes the time distance between the visits which is closely related to the frequency. The results reported in the paper are remarkable due to the extremely high Sp of 99.4%, but remakably enough the corresponding ER is as high as 93%. This illustrates the almost impossible task of achieving an overall satisfying performance of a detection model when using solely sensor-based data for detection of conditions with very low prevalence.

Kamphuis et al. (2010a,b) present decision trees with different data mining techniques or cost matrices added as improvements for detecting CM. Even though the inclusion of cost matrices in a model designed for decision support is highly relevant, it does not improve the performance enough for implementation.

5.3. Prioritizing methods

As seen in the descriptions of sheer and improved detection models, there is a general problem with fulfilling the described criteria for implementation. Scientific literature describes three overall alternative approaches to this problem; a higher extent of added knowledge, in the form of non-sensor information, to the original detection model (Fig. 3A), an acceptance of the original performance level plus a postprocessing step of prioritization or ranking of the alarms into TP or FP (Fig. 3B), or a presentation of the model output as a time gradient or a risk of case vs. non-case (Fig. 3C) disregarding the source of model input variables. To some extent, a customization of thresholds according to the risk attitude of the farmer can be regarded as a prioritizing measure, but this approach implies that the model is adjusted to the specific herd at time of implementation - and possibly multiple times hereafter as the health or managerial status is dynamic and will change.

Fuzzy logic is used by de Mol and Woldt (2001) to combine sensor-based output from earlier developed detection models with additional information about the cow (Fig. 3A) in order to formalize the manager's reasoning when manually judging alert lists for CM and oestrus. Hereby, they reduce the number of FP on the

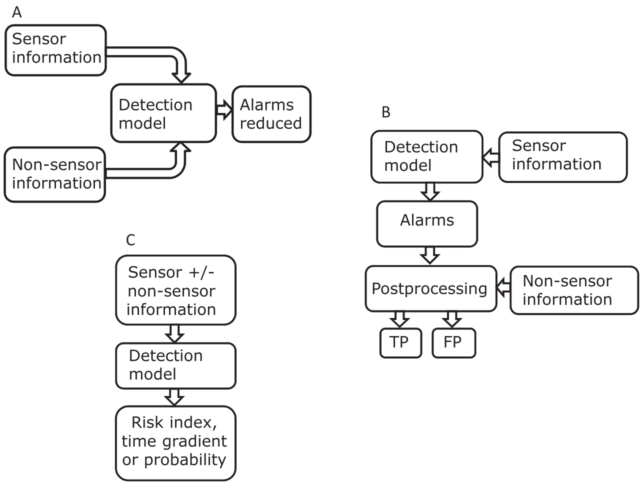


Fig. 3. Flow charts illustrating three prioritizing strategies for reducing the number of false alarms. Flow chart A illustrates a reduction in the number of generated alarms by using both sensor and non-sensor information as model input compared to using only sensor information as model input. Flow chart B illustrates a postprocess of the TP alarms generated by the detection model by adding, e.g. non-sensor information and hereby sorting the alarms into TP and FP. Flow chart C illustrates the cases of model output presented as a risk, a gradient up to or a probability of a condition happening.

CM alert list from 1265 to 64 and the number of oestrus alarms by 32%.

The CM model combines the AMS alerts from de Mol and Ouweltjes (2001) with average and variance of sensor measurements, while the oestrus model combines alerts from de Mol et al. (1997) with both qualitative and quantitative non-sensor-based cow information. By combining qualitative and quantitative parameters, de Mol and Woldt (2001) are fulfilling the basic demands of the Fuzzy logic method, but as remarkable as their results are, they must be interpreted with some care, since the Se and Sp respectively are calculated on different divisions of the data.

Naive Bayesian Networks (NBN) as a tool for discriminating between TP and FP alerts from AMS (Fig. 3B) is demonstrated in a study by Steeneveld et al. (2010a) where the number of FP alerts are reduced by 35%. Unfortunately, the model misses 10% of the TP alerts meaning that the specificity is too low for implementation. A satisfying performance level cannot be expected in this study though since the initial performance of the AMS providing the alert list has an Se of 70% and an Sp of 97.8%. The results do show a potential for NBN as a prioritizing method and more research should be done using this method.

A completely different approach for detecting or predicting a condition is used by Aparna et al. (2014). The paper focuses on predicting the exact onset of farrowing, in order to reduce piglet mortality caused by hypothermia. The underlying model is based on Hidden Phase-type Markov methodology, where the time spent in each defined phase of a given condition is modeled. For this study the well-defined behavioral phases preceding a farrowing is used.

The study is based on sows already inserted into farrowing section which makes the probability of the sow actually farrowing very high - almost definitely known to happen - unlike any other condition included in this review. Well-defined behavioral phases are known for a few conditions like parturition and to some extent oestrus in both sows and cows.

This phase-based method is, however, difficult to apply on conditions like CM or tail biting, where the chronological succession of phases is unknown, and different phase-patterns can lead to the same condition. Also a crucial difference between predicting the onset of farrowing and predicting events of CM and tail biting, is not knowing beforehand, whether the condition will occur at all or not.

Interestingly enough, Aparna et al. (2014) do not operate with the traditional performance parameters (sensitivity, specificity and error rate) but produce estimates of time to occurrence of farrowing (Fig. 3C) hereby providing decision support to the farmer in choosing which sow to attend to first. By combining water and activity sensors, the model produced 97% true warnings with a mean of 11.5 h and an SD of 4.6 h. These results fulfill the aim of the paper to provide sufficient warning time for preparing the crate and sow for farrowing, but the SD is too long to provide accurate alarms for the exact onset of the farrowing with the purpose of providing timely aid to complications.

6. Method discussion

The previous description of sheer, improved and prioritizing detection models illustrates a trend in the development of detection models over the last two decades (1995–2015). This trend is not depicting a straight forward evolution of detection-methods but rather a correlated evolution in both model complexity and general evolution of sensor technology.

6.1. Evolutionary trends of methods and sensors

The evolution of sensor-based detection models is facilitated both by the technological evolution causing lower market prices and smaller, more precise devices in general, and by the joint scientific experiences made through peer-reviewed studies and research. In that sense, the evolution of sensor-based detection models has generally moved from univariate models on general species level (Deluyker et al., 1990) or comparing data with a simple threshold (Bressers et al., 1995) through improving detection accuracy by including non-sensor-based animal-specific information like “day of treatment”(Cavero et al., 2007), “calving dates”, or “days in lactation” (de Mol et al., 2001). Parallel to including non-sensor-based information, more multivariate models were developed (Cavero et al., 2006; Kamphuis et al., 2008b; Kramer et al., 2009).

With performance still not reaching a satisfying level, researchers have continued to develop models focusing on prioritizing the generated alarms through the use of Fuzzy logic (de Mol and Woldt, 2001), Naive Bayesian Network (Steeneveld et al., 2010a) or variations of DLM (Cornou et al., 2008; Ostersen et al., 2010; de Mol et al., 2013). During the same period in time (1995–2015), the technical evolution of sensors has made it possible for the precision in CM detection to move from udder level to quarter level. Also data is available much faster, going from on-line monthly or weekly pooled reference data like SCC, to in-line sensors (Kamphuis et al., 2008b) providing the possibility of detecting a CM case during the actual milking.

A similar evolution of both sensor types and method complexity is also found in models detecting conditions like lameness in cows and oestrus in both sows and cows, but since the history of automatic detection is shorter for these conditions compared to CM, the evolutionary changes are not as profound.

Rajkondawar et al. (2002) developed a fully automatic detection model using limb-specific kinetic measures, and later several studies were based on partly automatic measures, using manual gait score as gold standard (Pastell and Kujala, 2007; Maertens et al., 2011). It has, however, been the development of force load cells (Liu et al., 2009) and pressure sensitive mats (Maertens et al., 2011), which has made a huge difference in lameness detection for cows. The former of the two sensor types, has even been used

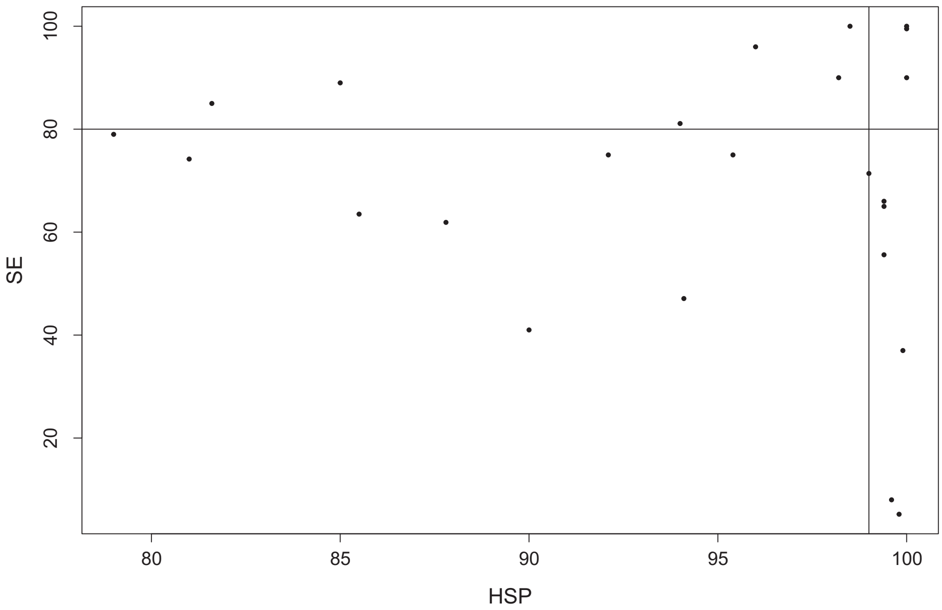


Fig. 4. Performance (Highest Specificity (HSp), corresponding specificity (Se)) for the 26 papers that present model performance with sensitivity and corresponding specificity. Lines indicate performance criteria (sensitivity 80% and specificity 99%).

in a commercially implemented product for lameness detection, which Liu et al. (2009) sought to make more accurate in their study.

6.2. The perfect performance - does it exist?

Despite the technological evolution and the increased complexity of methods in sensor-based detection models, the accuracies of these models are generally at a level that does not fulfill the criteria of implementation (Hogeveen et al., 2010). A great variation in model performance throughout the different studies is revealed when the performance is visualized. The performance of HSp and the corresponding Se is shown in Fig. 4 for the papers that present both Se and Sp. Nine studies reach a Se above 80% and ten reach a HSp above 99% but only three papers (Liu et al., 2009; Maertens et al., 2011; Cornou and Lundbye-Christensen, 2011) present models that fulfill both performance criteria when including subgroups of the data sets.

Liu et al. (2009) detect lameness in cows and use logistic regression in combination with B-spline transformation to obtain Se of 100 % and Sp of 100% when detecting lameness on forelimbs, and Se of 99.5% and Sp of 100% when detecting lameness at cow-level. The authors convert a five-point lameness score to a binary (sound-lame) and furthermore validate the model by the leave-one-out cross validation method on a data set consisting of 261 cows. It is reasonable to assume that leaving out the information provided by only one cow for validation, using the remaining 260 cows to train the model a total of 261 times, is close to learning and testing the model on the same data which will result in a high level of performance. Therefore, the study does not fulfill the similarity criteria. Maertens et al. (2011) obtain Se of 90% and Sp of 100% when detecting severe lame cows (gait score 3 on a three-point lameness score) using linear regression on kinematic variables from pressure sensitive mats, but the aim of lameness detection is primarily to point the farmer towards the cow that needs extra focus rather than those who need acute treatment (Pastell and Kujala, 2007), and with this model not fulfilling the performance-criteria for sound or mild-lame cows, it seems to be of little use in the production. Cornou and Lundbye-Christensen (2011) use CUSUM to detect the onset of farrowing based on the sow's activity pattern and obtain Se of 100% and Sp of 100% for a subgroup of 9 sows with the sow's individual variance. The level of performance when including all sows ($n = 19$) using individual variance is Se of 100% and Sp of 95% whereas the performance for all sows using group variance is Se of 95% and Sp of 89% thus not fulfilling the performance criteria. In the discussion, the authors mention that using individual variance might be over optimizing the model since the reference days of each sow were known beforehand. They recommend the study to be repeated in a large scale experiment where this bias is avoided and suggest inclusion of more animals and different setup of time windows.

The potential of sheer and improved sensor-based detection models is well exploited and they generally do not detect at an implementable level of accuracy. This calls for alternative approaches with a higher degree of customization and adaptability to individual needs at herd-, farmer-, or animal-levels.

6.3. Customization and prioritizing strategies

Throughout the literature, three strategies of prioritizing methods and few concrete suggestions for customizing models are described. Fig. 3A–C illustrate three different strategies for improving the performance of a model or for ranking or prioritizing the output of the detection models.

6.3.1. Customization

Customization of detection models based on DLM is suggested by Cornou et al. (2008), Ostersen et al. (2010), and de Mol et al. (2013) who all present different variations of DLM in their detection models and discuss further adaptation for implementation.

The DLM is not a prioritizing method as the previously discussed fuzzy logic and NBN in the manner of ranking alerts according to a given preference or classifying alerts as true or false. The DLM as a statistical method can predict - or produce a forecast for - the state of the condition of interest one step forward and compare the prediction with the following observation. Nevertheless, the method as presented in these three papers is capable of adjusting to individual circumstances through described strategies for changes in the discount factor of the DLM which alter the adaptability of the model.

The herd-specific adjustments could be on the prevalence of the condition in focus (Hogeveen et al., 2010), the level of management (Huijps et al., 2010) and the farmer's risk attitude. The latter might differ in terms of both economic consequences (Rutten et al., 2014) and workload associated with accepting a lower, or a higher, level of false alarms (Mollenhorst et al., 2012). In addition to this strategy de Mol et al. (2013) describe how changing the threshold for the Bayes factor of the DLM influences the Se and Sp of the model so it can be adjusted to the risk attitude or level of management at the individual herd.

6.3.2. Prioritizing strategies

Prioritizing strategy (A) combines sensor data with additional non-sensor information at animal-, section- or herd-specific level in a detection model in order to increase the level of performance. This strategy is followed to some extent by Maatje et al. (1997); Ostersen et al. (2010) who mention the potential of combining sensor and non-sensor data and by Garcia et al. (2014) where the parity of the cow is used as classification parameter when defining groups in the data set. Different methods can be used for combining sensor data and non-sensor data, and NBN has been used with interesting results in studies by Steeneveld et al. (2009, 2010b) who use cow-specific information to provide probability distributions for pathogens and for prioritizing alerts from AMS alert lists.

NBN is also used by Jensen et al. (2016) for combining sensor data and cow-specific information in a CM detection model. Using NBN for combining data from different sources is not common in livestock production but has been done previously by Steeneveld et al. (2009, 2010b) and also recently in the world of computer security where Benferhat et al. (2013); Bouzar-Benlabiod et al. (2015) combine sensor alerts and expert knowledge to improve performance of computer security models.

Even though animal-specific information has great impact on the performance of a detection model, the use of cow-specific information alone is not always enough as proven by Steeneveld et al. (2008, 2010a). Animal-specific biological markers, as used by Chagunda et al. (2006) in a dynamic deterministic biological model, can however show that detailed cow-specific information in combination with laboratory analysis of the enzyme L-lactate dehydrogenase (LDH) can present an impressive performance level with Se 82% and Sp 99% - including no other AMS information. This type of model is yet not implementable due to technological demands.

Prioritizing strategy (B) describes a different approach where the sub-optimal performance obtained by a detection model, whether based either solely on sensors or on combined information types, is initially accepted and the generated alarms are prioritized or ranged by combining them with additional non-sensor data in a following postprocessing step as it has been done by de Mol and Woldt (2001); Steeneveld et al. (2010a) (see Fig. Flow chart B 1,2). Two different methods using strategy (B) are described in the literature; fuzzy logic and NBN.

By using fuzzy logic de Mol and Woldt (2001) reduce the number of false positive alerts from earlier developed statistical models detecting CM and/or oestrus (de Mol et al., 1997, CM and oestrus); (de Mol and Ouweltjes, 2001, CM). Two separate fuzzy logic models are created - one for each condition. In the CM model de Mol and Woldt (2001) reduce the number of false positive alarms from 1265 to 64 by combining the output of the statistical model with fuzzified additional information. The information is added on standardized deviation in electric conductivity of each quarter as well as measured conductivity at quarter level. This use of fuzzy logic raises the same issue as seen in Cavero et al. (2006), Kamphuis et al. (2008b), Kramer et al. (2009) where numerical values are first fuzzified to lingual values, and then defuzzified to numerical again. Since fuzzy logic is a method meant for quantifying linguistic - or fuzzy - values, it seems more obvious to use the method on qualitative factors like reproductive status, level of activity or a description of lameness degree parallel to lameness scores in lameness detection.

Furthermore, the input to this fuzzy logic model is the output of a statistical model where the performance is obtained through long time windows and a high degree of selectivity in the choice of included data (de Mol and Ouweltjes, 2001). As opposed to the CM model, the oestrus model include qualitative parameters like reproductive status and information on activity level. The number of false positive alerts were reduced by 32%, and the false alarms were sought to be further reduced through manual and computational optimization of the model but without noteworthy improvements. In their discussion the authors discuss that the model might have been improved further by including the use of "expert knowledge" from the herdsman or personnel. Even though de Mol and Woldt (2001) reduce the number of false positive alarms, and present a method of prioritization, it is our opinion that fuzzy logic should be used with care on data sets consisting of large amount of quantitative information like sensor-based data as the method is not well suited for this.

Steeneveld et al. (2010a) also follow strategy (B) and use Naive Bayesian Network (NBN) to classify which of the alerted cows on an AMS alert list need further investigation for CM. This is done by calculating the probability of an alert being TP or TN based on information from either one variable or combinations of variables. The variables originate either solely from AMS, solely from additional cow-specific information, or from combining this information. The AUC clearly shows that combining the two sources of information perform the best. NBN is well suited for expressing uncertainties, which will inevitable be a part of describing large individual variation. Even though NBN is the simplest version of Bayesian Classification models, assuming no correlation between the included variables, more advanced Bayesian Networks have been tested on the same data sets without improving the results (Steeneveld et al., 2010a).

Interestingly enough when analyzing the impact of single variables, Steeneveld et al. (2010a) find that from the non-AMS cow information (parity, days in milk, season of year, SCC history, CM alert history) only days in milk were significantly different between FP and TP alerts. On the opposite, high levels of SCC found in the SCC history of the cow were evenly distributed among the cows generating FP and TP alerts. Because the level of SCC is considered a very important indicator of CM (Steeneveld et al., 2008), and the SCC level measured by Steeneveld et al. (2010a) was significantly different between TP and TN milkings, these findings indicate that the alert list is based solely on SCC. This indication makes the ranking of alarms based on multiple cow-specific parameters - not only on SCC - highly relevant. High levels of SCC provide valuable information, but just generate too many FP, perhaps even detecting both CM and subclinical CM when used as single variable (Rasmussen and Bjerring, 2005; Steeneveld et al., 2010a).

Steeneveld et al. (2010a) do not reach a satisfying accuracy when discriminating between TP and FP alerts, but the number of FP is reduced by 33%, and the use of NBN as a simple prioritizing tool in livestock production herds warrants further consideration. The capability of NBN to combine information through adding prior probabilities for any relevant information, sensor-based or not, enables the incorporation of managerial factors. These factors could be changes in feed composition, treatments, and herd-specific routines. Information on the herd-specific health status is also relevant for evaluating if the conditions of interest is of higher or lower prevalence than in average herds.

An important aspect in customizing an early warning system to a specific herd or risk-attitude of a farmer is the farmer's preferences to the detection system. Mollenhorst et al. (2012) have asked farmers what preferences they have to a CM detection system, and the result is that a low number of false alerts and alerts given in good time with emphasis on the more severe cases is the most important feature. The adaptability to individual circumstances is also important for the farmers according to the questionnaire by Mollenhorst et al. (2012). The probabilities for any relevant information can be combined with herd-specific thresholds according to the priorities and risk attitude of the farmer (Steeneveld et al., 2010a). A method like NBN shows this high degree of adaptation, hence meeting the demands for customization, characteristic for modern farmers with ambitions (Mollenhorst et al., 2012).

Prioritizing strategy (C) represents an alternative to performance presented by the epidemiological terms of Se and Sp. This alternative is to present the output of the detection model as a gradient or a risk of a condition occurring. Se and Sp are designed for binary outputs, which essentially does not conditions like CM, tail biting, or lameness, which are gradually evolving, and in nature more complex than binary (Friggens et al., 2007, 2010). Detection models in livestock production are, however, traditionally based on discrete measurements in time (Sherlock et al., 2008) which simplify the picture of a complex condition. Presenting the alarms in the form of a risk indicator (Nielsen et al., 2005; Friggens et al., 2007; Hojsgaard and Friggens, 2010) or as a time gradient leading up to the occurrence of a condition known to happen (Aparna et al., 2014) has been seen. In addition to these output types, the posterior probabilities for a condition to occur as calculated by NBN could be a future approach worth focusing on. Strategy (C) is well suited as a decision support tool because it provides detailed information on the individual animal and at the same time allows the farmer to evaluate the alarms personally and use both experience and knowledge of the herd in combination with well substantiated information from the detection model.

6.4. What is more important - priorities are dynamic

In this review, the overall perspective for evaluating the prioritizing detection models has been to reduce the number of false alarms communicated to the farmer. Traditionally the models have generated alarms indicating what animal to attend to, but other motivations for prioritizing can be mentioned. Decision support for which intervention to choose if multiple are possible, or which alarms to attend to first if more monitoring systems are installed at the same farm generating alarms at the same time are both relevant. The optimal prioritization is not a static solution. It might change on a weekly or even daily basis according to multiple factors, and different interests could generate different optimal prioritization outputs. Market prices or costs associated with an intervention (man-hours, equipment, etc.) could be used as added information parameters in a prioritizing model. Such a cost minimizing approach would most likely generate a different output than using animal health parameters or welfare parameters. From the farmers perspective, it might be of high priority to optimize his or hers life quality by generating more free time to spend with the family or by increasing the social acceptance in society.

6.5. Research perspectives for early warning systems

The field of automatic monitoring and modeling is still relatively young, and concurrently with the technological evolution, future perspectives for developing decision-supporting tools for ambitious livestock producers continue to be an extremely interesting field of research and development. This review only includes papers that present a concrete performance, but many studies are exploring a range of topics, including lameness detection (Rajkondawar et al., 2002, 2006; Pastell et al., 2008a,b; Pastell and Madsen, 2008; XiangYu et al., 2008; Chapinal et al., 2009; Nielsen et al., 2010; Tanida et al., 2011; Pluk et al., 2012; Hoffmann et al., 2013; Van Nuffel et al., 2013; Pluym et al., 2013; Abell et al., 2014; Hothersall et al., 2014; Mohling et al., 2014; Wood et al., 2015), vision-based monitoring (White et al., 2004; Porto et al., 2014; Leroy et al., 2008; Cangar et al., 2008; XiangYu et al., 2008; Abdanan Mehdizadeh et al., 2015; Kristensen and Cornou, 2011; Kashiha et al., 2013), methods for reducing animal mortality (Beltrán-Alcrudo et al., 2009; Bono et al., 2012, 2013, 2014), modeling of behavioral traits as welfare indicators (Bressers et al., 1994; Turner et al., 2000; Moshou et al., 2001; Madsen et al., 2005a; Madsen and Kristensen, 2005; Oliviero et al., 2008; Ferrari et al., 2010; Junge et al., 2012; Cornou and Kristensen, 2014a; Dutta et al., 2015), as well as the continuing focus on detecting CM in dairy cows (Kamphuis et al., 2008a; Claycomb et al., 2009; Lukas et al., 2009).

Sceptics might argue that further research in the development of early warning systems is of little use since the criteria for implementation are so difficult to fulfill. But looking at the broader perspectives, automatic monitoring and early warning systems offer an opportunity to observe the animals 24 h a day 7 days week 365 days a year, which is far more than what is human possible in traditional livestock production. Early warning systems will always be a decision support tool for the farmer, and not a bullet-proof management manual. The farmer accepting a certain amount of false alarms, or relating to a given risk indicator for a condition occurring, is a realistic scenario after the implementation of a sensor-based early warning system. The perspectives for improving animal welfare through precision livestock farming are distinct, although more research is needed before warning systems with sufficient accuracy are ready for implementation.

7. Conclusion

Three methods have been used for prioritizing sensor-based alarms in livestock production. Two of these methods, Fuzzy logic and Naive Bayesian Network, combine sensor data with non-sensor data whereas the third method, Hidden phase-type Markov model, generates a time gradient to the onset of farrowing - a condition known to happen. The use of Fuzzy logic reduces the number of alarms considerably but the method is not well suited for data consisting of large amounts of numerical values like sensor-based data.

Naive Bayesian Network reduces the number of alarms by 57%, and this method shows potential for further research in prioritizing true and false alarms. Hidden phase-type Markov model generates a continuous output which is an interesting alternative to the binary Se and Sp although the Hidden phase-type Markov model might not be the right choice for modeling conditions with no - or diffusely defined - phases or with varying probabilities of occurrence.

For 20 years, no sensor-based detection model has fulfilled the performance demands needed to generate a satisfyingly low level of false positive alarms, and these demands seem close to unreachable with the few models actually obtaining high performances being associated with high error rates. Instead of focusing on fulfilling unreachable demands based on binary performance parameters for more complex conditions, future research could seek alternative approaches for the output of detection models as for instance the prior probability - or the risk of a condition occurring or not. Alarms from detection models can be prioritized in order to optimize production efficiency, production costs, work load and animal health, and a future with automatic monitoring in livestock production looks promising considering both the life quality of the farmer and the welfare of the animals.

References

Abell, C.E., Johnson, A.K., Karriker, L.A., Rothschild, M.F., Hoff, S.J., Sun, G., Fitzgerald, R.F., Stalder, K.J., 2014. Using classification trees to detect induced sow lameness with a transient model. Animal 8 (6), 1000–1009. Update Code: 20140625.

Alsaaod, M., Romer, C., Kleinmanns, J., Hendriksen, K., RoseMeierhofer, S., Plumer, L., Buscher, W., 2012. Electronic detection of lameness in dairy cows through measuring pedometric activity and lying behavior. Appl. Anim. Behav. Sci. 142 (3/4), 134–141. Update Code: 20130130.

Aparna, U., Pedersen, L.J., Jorgensen, E., 2014. Hidden phase-type markov model for the prediction of onset of farrowing for loose-housed sows. Comp. Electron. Agricult. 108, 135–147 (Update Code: 20141203).

Barkema, H.W., Schukken, Y.H., Lam, T.J.G.M., Beiboer, M.L., Wilmink, H., Benedictus, G., Brand, A., 1998. Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. J. Dairy Sci. 81 (2), 411–419.

Beltrán-Alcrudo, Daniel, Carpenter, Tim E., Cardona, Carol, 2009. A flock-tailored early warning system for low pathogenic avian influenza (LPAI) in commercial egg laying flocks. Prevent. Veter. Med. 92 (4), 324–332. 12/1.

Benferhat, Salem, Boudjelida, Abdelhamid, Tabia, Karim, Drias, Habiba, 2013. An intrusion detection and alert correlation approach based on revising probabilistic classifiers using expert knowledge. Appl. Intell. 38 (4), 520–540 (JUN. PT: J; NR: 70; TC: 4; J9: APPL INTELL; PG: 21; GA: 140IE; UT: WOS:000318646900005).

Berckmans, D., 2014. Precision livestock farming technologies for welfare management in intensive livestock systems (special issue: animal welfare: focusing on the future). Revue Scientifique et Technique - Office International des Epizooties 33 (1), 189–196 (ID: 105; Update Code: 20140730).

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. Livest. Sci. 149 (3), 289–300. Update Code: 20121205.

Bono, C., Cornou, C., LundbyeChristensen, S., Kristensen, A.R., 2013. Dynamic production monitoring in pig herds ii. Modeling and monitoring farrowing rate at herd level. Livest. Sci. 155 (1), 92–102 (Update Code: 20130731).

Bono, C., Cornou, C., LundbyeChristensen, S., Kristensen, A.R., 2014. Dynamic production monitoring in pig herds iii. modeling and monitoring mortality rate at herd level. Livest. Sci. 168, 128–138 (Update Code: 20141119).

Bouzar-Benlabiod, Lydia, Benferhat, Salem, Bouabana-Tebibel, Thouraya, 2015. Instantiated first order qualitative choice logic for an efficient handling of alerts correlation. Intell. Data Anal. 19 (1), 3–27 (PT: J; NR: 33; TC: 0; J9: INTELL DATA ANAL; PG: 25; GA: AW7AW; UT: WOS:000346417800002).

Bressers, H.P.M., te Brake, J.H.A., Jansen, M.B., Nijenhuis, P.J., Noordhuizen, J.P.T.M., 1994. Monitoring individual sows: radiotelemetrically recorded ear base temperature changes around farrowing. Livest. Product. Sci. 37 (3), 353–361 (Update Code: 19940000).

Bressers, H.P.M., Te Brake, J.H.A., Noordhuizen, J.P.T.M., 1995. Automated oestrus detection in group-housed sows by recording visits to the boar. Livest. Product. Sci. 41 (2), 183–191.

Cangar, O., Leroy, T., Guarino, M., Vranken, E., Fallon, R., Lenehan, J., Mee, J., Berckmans, D., 2008. Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis (special issue: Smart sensors in precision livestock farming). Comp. Electron. Agricult. 64 (1), 53–60 (Update Code: 20080000).

Cavero, D., Tolle, K.H., Buxade, C., Krieter, J., 2006. Mastitis detection in dairy cows by application of fuzzy logic. Livest. Sci. 105 (1/3), 207–213 (ID: 41; Update Code: 20060000).

Cavero, D., Tolle, K.H., Rave, G., Buxade, C., Krieter, J., 2007. Analysing serial data for mastitis detection by means of local regression. Livest. Sci. 110 (1/2), 101–110 (Update Code: 20070000).

Chagunda, M.G.G., Friggens, N.C., Rasmussen, M.D., Larsen, T., 2006. A model for detection of individual cow mastitis based on an indicator measured in milk. J. Dairy Sci. 89 (8), 2980–2998 (CAS Registry Numbers: 9001-60-9; Update Code: 20060000).

Chapinal, N., de PassillT, A.M., Weary, D.M., von Keyserlingk, M.A.G., Rushen, J., 2009. Using gait score, walking speed, and lying behavior to detect hoof lesions in dairy cows. J. Dairy Sci. 92 (9), 4365–4374.

Claycomb, R.W., Johnstone, P.T., Mein, G.A., Sherlock, R.A., 2009. An automated in-line clinical mastitis detection system using measurement of conductivity from foremilk of individual udder quarters. New Zeal. Veterin. J. 57 (4), 208–214 (Update Code: 20090000).

Cornou, C., Kristensen, A.R., 2014a. Monitoring individual activity before, during and after parturition using sensors for sows with and without straw amendment. Livest. Sci. 168, 139–148 (Update Code: 20141119).

Cornou, C., Kristensen, A.R., 2014b. Monitoring individual activity before, during and after parturition using sensors for sows with and without straw amendment. Livest. Sci. 168, 139–148 (Update Code: 20141119).

Cornou, C., Lundbye-Christensen, S., 2010. Classification of sows' activity types from acceleration patterns using univariate and multivariate models. Comp. Electron. Agricult. 72 (2), 53–60 (Update Code: 20100000).

Cornou, C., Lundbye-Christensen, S., 2011. Modeling of sows diurnal activity pattern and detection of parturition using acceleration measurements. Comp. Electron. Agricult. 80, 97–104 (Update Code: 20120000).

Cornou, C., Vinther, J., Kristensen, A.R., 2008. Automatic detection of oestrus and health disorders using data from electronic sow feeders. Livest. Sci. 118 (3), 262–271 (Update Code: 20080000).

Cornou, C., LundbyeChristensen, S., Kristensen, A.R., 2011. Modelling and monitoring sows' activity types in farrowing house using acceleration data. Comp. Electron. Agricult. 76 (2), 316–324 (Update Code: 20110000).

de Mol, R.M., Ouweltjes, W., 2001. Detection model for mastitis in cows milked in an automatic milking system. Prevent. Veter. Med. 49 (1/2), 71–82 (ID: 53; Update Code: 20010000).

de Mol, R.M., Woldt, W.E., 2001. Application of fuzzy logic in automated cow status monitoring. J. Dairy Sci., 84(2) (ID: 51; Note: Includes references; Identifier: false positive results. diagnostic value.; Category Code: Veterinary Science [L800]. Mathematics And Statistics [X100].; NAL Location: 44.8 J822.; Update Code: 200505).

de Mol, R.M., Kroeze, G.H., Achten, J.M.F.H., Maatje, K., Rossing, W., 1997. Results of a multivariate approach to automated oestrus and mastitis detection. Livest. Product. Sci. 48 (3), 219–227. 6.

de Mol, R.M., Keen, A., Kroeze, G.H., Achten, J.M.F.H., 1999. Description of a detection model for oestrus and diseases in dairy cattle based on time series analysis combined with a kalman filter. Comp. Electron. Agricult. 22(2/3), 171–185 (16–19 June, 1996 1999. Update Code: 19990000).

de Mol, R.M., Ouweltjes, W., Kroeze, G.H., Hendriks, M.M.W.B., 2001. Detection of oestrus and mastitis: field performance of a model. Appl. Eng. Agricult. 17 (3), 399–407 (Update Code: 20020000).

de Mol, R.M., Andre, G., Bleumer, E.J.B., van der Werf, J.T.N., de Haas, Y., Reenen, C.G., 2013. Applicability of day-to-day variation in behavior for the automated detection of lameness in dairy cows. J. Dairy Sci. 96 (6), 3703–3712 (Update Code: 20130619).

Deluyker, H.A., Shumway, R.H., Wecker, W.E., Azari, A.S., Weaver, L.D., 1990. Modeling daily milk yield in holstein cows using time series analysis. J. Dairy Sci. 73 (2), 539–548 (Update Code: 19900000).

Dutta, R., Smith, D., Rawnsley, R., BishopHurley, G., Hills, J., Timms, G., Henry, D., 2015. Dynamic cattle behavioural classification using supervised ensemble classifiers. Comp. Electron. Agricult. 111, 18–28 (Update Code: 20150415).

Ferrari, S., Piccinini, R., Silva, M., Exadaktylos, V., Berckmans, D., Guarino, M., 2010. Cough sound description in relation to respiratory diseases in dairy calves. Prevent. Veter. Med. 96 (3/4), 276–280 (Update Code: 20100000).

Firk, R., Stamer, E., Junge, W., Krieter, J., 2002. Automation of oestrus detection in dairy cows: a review. Livest. Product. Sci. 75 (3), 219–232. 7.

Freson, L., Godrie, S., Bos, N., Jourquin, J., Geers, R., 1998. Validation of an infra-red sensor for oestrus detection of individually housed sows. Comp. Electron. Agricult. 20 (1), 21–29 (Update Code: 19980000).

Friggens, N.C., Chagunda, M.G.G., Bjerring, M., Ridder, C., Hojsgaard, S., Larsen, T., 2007. Estimating degree of mastitis from time-series measurements in milk: a test of a model based on lactate dehydrogenase measurements. J. Dairy Sci. 90 (12), 5415–5427 (ID: 79; CAS Registry Numbers: 9001-60-9; Update Code: 20080000).

Friggens, N.C., Codrea, M.C., Hojsgaard, S., 2010. Extracting biologically meaningful features from time-series measurements of individual animals: towards quantitative description of animal status. In: Sauvant, D., van Milgen, J., Faverdin, P., Friggens, N. (Eds.), 7th International Workshop on Modelling Nutrient Digestion and Utilisation in Farm Animals, Wageningen, 10–12 September, 2009 (Faculty of Agricultural Sciences, University of Aarhus, P.O. Box 50, 8830 Tjele, Denmark.; nicolas.friggens@agroparistech.fr, Wageningen Academic Publishers. ISBN 978-90-8686-156-9. 40-48; Update Code: 20110000).

Garcia, E., Klaas, I., Amigo, J.M., Bro, R., Enevoldsen, C., 2014. Lameness detection challenges in automated milking systems addressed with partial least squares discriminant analysis. J. Dairy Sci. 97 (12), 7476–7486 (Update Code: 20150107).

Hoffmann, G., Ammon, C., Volkamer, L., Surie, C., Radko, D., 2013. Sensor-based monitoring of the prevalence and severity of foot pad dermatitis in broiler chickens. Brit. Poul. Sci. 54 (5), 553–561 (Update Code: 20140108).

Hogeveen, H., Kamphuis, C., Steeneveld, W., Mollenhorst, H., 2010. Sensors and clinical mastitis - the quest for the perfect alert. Sensors 10 (9), 7991–8009 (Update Code: 20100000).

Hojsgaard, S., Friggens, N.C., 2010. Quantifying degree of mastitis from common trends in a panel of indicators for mastitis in dairy cows. J. Dairy Sci. 93 (2), 582–592 (ID: 80; Update Code: 20100000).

Hothersall, B., Caplen, G., Parker, R.M.A., Nicol, C.J., WatermanPearson, A.E., Weeks, C.A., Murrell, J.C., 2014. Thermal nociceptive threshold testing detects altered sensory processing in broiler chickens with spontaneous lameness. PLoS One 9 (5) (Update Code: 20140730).

Huijps, K., Hogeveen, H., Antonides, G., Valeeva, N.I., Lam, T.J.G.M., Lansink, A.G.J.M. O., 2010. Sub-optimal economic behaviour with respect to mastitis management. Euro. Rev. Agricult. Econ. 37 (4), 553–568 (ID: 64; Update Code: 20100000).

Huybrechts, T., Mertens, K., De Baerdemaeker, J., De Ketelaere, B., Saeys, W., 2014. Early warnings from automatic milk yield monitoring with online synergistic control. J. Dairy Sci. 97 (6), 3371–3381. 6.

Jensen, Dan B., Hogeveen, Henk, Vries, Albert De, 2016. Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis. J. Dairy Sci. 99 (9), 7344–7361. ISSN 0022-0302.

Junge, M., Herd, D., Jezierny, D., Gallmann, E., Jungbluth, T., 2012. Water intake and drinking behavior of pregnant sows. In: The Ninth International Livestock Environment Symposium (ILES IX). International Conference of Agricultural Engineering - CIGR-AgEng 2012: Agriculture and Engineering for a Healthier Life, Valencia, 8–12 July 2012 (Livestock Systems Engineering, Institute of

Agricultural Engineering, University of Hohenheim, Garbenstrasse 9, 70593 Stuttgart, Germany.; mjunge@uni-hohenheim.de, C-; CIGR-EurAgEng. Update Code: 20131030).
Kamphuis, C., Pietersma, D., van der Tol, R., Wiedemann, M., Hogeveen, H., 2008a. Using sensor data patterns from an automatic milking system to develop predictive variables for classifying clinical mastitis and abnormal milk. Comp. Electron. Agricult. 62 (2), 169–181 (ID: 67; Update Code: 20080000).
Kamphuis, C., Sherlock, R., Jago, J., Mein, G., Hogeveen, H., 2008b. Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. J. Dairy Sci. 91 (12), 4560–4570 (ID: 66; Update Code: 20090000).
Kamphuis, C., Mollenhorst, H., Feelders, A., Pietersma, D., Hogeveen, H., 2010a. Decision-tree induction to detect clinical mastitis with automatic milking. Comp. Electron. Agricult. 70 (1), 60–68 (ID: 52; Update Code: 20100000).
Kamphuis, C., Mollenhorst, H., Heesterbeek, J.A.P., Hogeveen, H., 2010b. Detection of clinical mastitis with sensor data from automatic milking systems is improved by using decision-tree induction. J. Dairy Sci. 93 (8), 3616–3627 (ID: 65; Update Code: 20100000).
Kamphuis, C., Frank, E., Burke, J.K., Verkerk, G.A., Jago, J.G., 2013. Applying additive logistic regression to data derived from sensors monitoring behavioral and physiological characteristics of dairy cows to detect lameness. J. Dairy Sci. 96 (11), 7043–7053 (Update Code: 20131120).
Kashiha, M., Bahr, C., Haredasht, S.A., Ott, S., Moons, C.P.H., Niewold, T.A., Odberg, F. O., Berckmans, D., 2013. The automatic monitoring of pigs water use by cameras. Comp. Electron. Agricult. 90, 164–169 (ID: 92; Update Code: 20130306).
Kashiha, M., Bahr, C., Ott, S., Moons, C.P.H., Niewold, T.A., Odberg, F.O., Berckmans, D., 2014. Automatic weight estimation of individual pigs using image analysis. Comp. Electron. Agricult. 107, 38–44 (Update Code: 20141001).
Klir, George J., Folger, Tina A., 1988. Fuzzy Sets, Uncertainty, and Information. Prentice Hall.
Kramer, E., Cavero, D., Stamer, E., Krieter, J., 2009. Mastitis and lameness detection in dairy cows by application of fuzzy logic. Livest. Sci. 125 (1), 92–96 (Update Code: 20090000).
Kristensen, A.R., Jørgensen, E., Toft, N., 2010. Herd Management Science. I. Basic concepts. Academic books, Copenhagen (ISBN 9788763461191).
Kristensen, Helle H., Cornou, CTcile, 2011. Automatic detection of deviations in activity levels in groups of broiler chickens a pilot study. Biosyst. Eng. 109 (4), 369–376. 8.
Leroy, T., Borgonovo, F., Costa, A., Aerts, J.M., Guarino, M., Berckmans, D., 2008. Real-time measurement of pig activity in practical conditions. In: Central Theme, Technology For All: Sharing the Knowledge for Development. Proceedings of the International Conference of Agricultural Engineering, XXXVII Brazilian Congress of Agricultural Engineering, International Livestock Environment Symposium - ILES VIII, Iguassu Falls City, Bonn, 31st August to 4th September, 2008 2008. International Commission of Agricultural Engineering (CIGR), Institut fur Landtechnik (ID: 116 Update Code: 20100000).
Liu, J., Neerchal, N.K., Tasch, U., Dyer, R.M., Rajkondawar, P.G., 2009. Enhancing the prediction accuracy of bovine lameness models through transformations of limb movement variables. J. Dairy Sci. 92 (6), 2539–2550. 6.
Lukas, J.M., Reneau, J.K., Wallace, R., Hawkins, D., Munoz-Zanzi, C., 2009. A novel method of analyzing daily milk production and electrical conductivity to predict disease onset. J. Dairy Sci. 92 (12), 5964–5976. 12.
Maatje, K., de Mol, R.M., Rossing, W., 1997. Cow status monitoring (health and oestrus) using detection sensors. Comp. Electron. Agricult. 16 (3), 245–254. 2.
Madsen, T.N., Kristensen, A.R., 2005. A model for monitoring the condition of young pigs by their drinking behaviour. Comp. Electron. Agricult. 48 (2), 138–154 (Update Code: 20050000).
Madsen, T.N., Andersen, S., Kristensen, A.R., 2005a. Modelling the drinking patterns of young pigs using a state space model. Comp. Electron. Agricult. 48 (1), 39–62 (Update Code: 20050000).
Madsen, T.N., Andersen, S., Kristensen, A.R., 2005b. Modelling the drinking patterns of young pigs using a state space model. Comp. Electron. Agricult. 48 (1), 39–62 (24 ref, Update Code: 20050000).
Maertens, Willem, Vangeyte, Jnrgen, Baert, Jeroen, Jantuan, Alexandru, Mertens, Koen C., De Campeneere, Sam, Pluk, Arno, Opsomer, Geert, Van Weyenberg, Stephanie, Van Nuffel, Annelies, 2011. Development of a real time cow gait tracking and analysing tool to assess lameness using a pressure sensitive walkway: the gaitwise system. Biosyst. Eng. 110 (1), 29–39. 9.
Abdanan Mehdizadeh, S., Neves, D.P., Tscharke, M., NSSs, I.A., Banhazi, T.M., 2015. Image analysis method to evaluate beak and head motion of broiler chickens during feeding. Comp. Electron. Agricult. 114 (0), 88–95. 6.
Meijer, E., Oosterlinck, M., van Nes, A., Back, W., Staay, F.J., 2014. Pressure mat analysis of naturally occurring lameness in young pigs after weaning. BMC Veter. Res. 10(193) (August 2014. Update Code: 20140917).
Mein, G.A., Rasmussen, M.D., 2008. Performance evaluation of systems for automated monitoring of udder health: would the real gold standard please stand up? In: Lam, T.J.G.M. (Ed.), Mastitis Control: From Science to Practice. Proceedings of International Conference, Wageningen, 30 September - 2 October 2008. Sensortec Ltd, Hamilton, New Zealand (mein@netspace.net.au, Mastitis control: from science to practice. Proceedings of International Conference, The Hague, Netherlands; Wageningen Academic Publishers. ISBN 978-90-8686-085-2. ID: 34; September - 2 October 2008; 2008.:259-266; Update Code: 20090000).
Miekley, B., Traulsen, I., Krieter, J., 2012. Detection of mastitis and lameness in dairy cows using wavelet analysis. Livest. Sci. 148 (3), 227–236 (Update Code: 20120926).
Miekley, B., Stamer, E., Traulsen, I., Krieter, J., 2013a. Implementation of multivariate cumulative sum control charts in mastitis and lameness monitoring. J. Dairy Sci. 96 (9), 5723–5733 (Update Code: 20130918).
Miekley, B., Traulsen, I., Krieter, J., 2013b. Principal component analysis for the early detection of mastitis and lameness in dairy cows. J. Dairy Res. 80 (3), 335–343 (Update Code: 20130821).
Milner, P., Page, K.L., Walton, A.W., Hillerton, J.E., 1996. Detection of clinical mastitis by changes in electrical conductivity of foremilk before visible changes in milk. J. Dairy Sci. 79 (1), 83–86 (Update Code: 19960000).
Mohling, C.M., Johnson, A.K., Coetzee, J.F., Karriker, L.A., Abell, C.E., Millman, S.T., Stalder, K.J., 2014. Kinematics as objective tools to evaluate lameness phases in multiparous sows. Livest. Sci. 165 (0), 120–128. 7.
Mollenhorst, H., Rijkaart, L.J., Hogeveen, H., 2012. Mastitis alert preferences of farmers milking with automatic milking systems. J. Dairy Sci. 95 (5), 2523–2530 (Update Code: 20120000).
Moshou, D., Chedad, A., van Hirtum, A., de Baerdemaeker, J., Berckmans, D., Ramon, H., 2001. An intelligent alarm for early detection of swine epidemics based on neural networks. Trans. ASAE 44(1) [Erratum: Mar/Apr 2001, v. 44 (2), p. 457.]; Note: Includes references; Identifier: intelligent systems.; Category Code: Veterinary Science [L800]. Animal Production [L100]. Computer And Library Sciences [X200].; NAL Location: 290.9 Am32T.; Update Code: 200505.
Nielsen, L.R., Pedersen, A.R., Herskin, M.S., Munksgaard, L., 2010. Quantifying walking and standing behaviour of dairy cows using a moving average based on output from an accelerometer. Appl. Anim. Behav. Sci. 127 (1/2), 12–19 (Update Code: 20100000).
Nielsen, N.I., Friggens, N.C., Chagunda, M.G.G., Ingvartsen, K.L., 2005. Predicting risk of ketosis in dairy cows using in-line measurements of beta-hydroxybutyrate: a biological model. J. Dairy Sci. 88 (7), 2441–2453 (CAS Registry Numbers: 300-85-6; Update Code: 20050000).
Van Nuffel, A., Vangeyte, J., Mertens, K.C., Pluym, L., De Campeneere, S., Saeys, W., Opsomer, G., Van Weyenberg, S., 2013. Exploration of measurement variation of gait variables for early lameness detection in cattle using the gaitwise. Livest. Sci. 156 (13), 88–95. 9.
Oliviero, C., Pastell, M., Heinonen, M., Heikkonen, J., Valros, A., Ahokas, J., Vainio, O., Peltoniemi, O.A.T., 2008. Using movement sensors to detect the onset of farrowing. Biosyst. Eng. 100 (2), 281–285 (Update Code: 20080000).
Ostersen, T., Cornou, C., Kristensen, A.R., 2010. Detecting oestrus by monitoring sows' visits to a boar. Comp. Electron. Agricult. 74 (1), 51–58. 10.
Pastell, M., Hautala, M., Poikalainen, V., Praks, J., VeermSe, I., Kujala, M., Ahokas, J., 2008a. Automatic observation of cow leg health using load sensors. Comp. Electron. Agricult. 62 (1), 48–53. 6.
Pastell, M., Kujala, M., Aisla, A.M., Hautala, M., Poikalainen, V., Praks, J., Veermae, I., Ahokas, J., 2008b. Detecting cow's lameness using force sensors. (special issue: Smart sensors in precision livestock farming.). Comp. Electron. Agricult. 64 (1), 34–38 (Update Code: 20080000).
Pastell, M.E., Kujala, M., 2007. A probabilistic neural network model for lameness detection. J. Dairy Sci. 90 (5), 2283–2292. 5.
Pastell, Matti, Madsen, Henrik, 2008. Application of cusum charts to detect lameness in a milking robot. Exp. Syst. Appl. 35 (4), 2032–2040. 11.
Pluk, A., Bahr, C., Poursaberi, A., Maertens, W., van Nuffel, A., Berckmans, D., 2012. Automatic measurement of touch and release angles of the fetlock joint for lameness detection in dairy cattle using vision techniques. J. Dairy Sci. 95 (4), 1738–1748. 4.
Pluym, Liesbet M., Maes, Dominiek, Vangeyte, Jnrgen, Mertens, Koen, Baert, Jeroen, Van Weyenberg, Stephanie, Millet, Sam, Van Nuffel, Annelies, 2013. Development of a system for automatic measurements of force and visual stance variables for objective lameness detection in sows: Sowsis. Biosyst. Eng. 116 (1), 64–74. 9.
Porto, S.M.C., Arcidiacono, C., Giummarra, A., Anguzza, U., Cascone, G., 2014. Localisation and identification performances of a real-time location system based on ultra wide band technology for monitoring and tracking dairy cow behaviour in a semi-open free-stall barn. Comp. Electron. Agricult. 108 (0), 221–229. 10.
Quimby, W.F., Sowell, B.F., Bowman, J.G.P., Branine, M.E., Hubbert, M.E., Sherwood, H.W., 2001. Application of feeding behaviour to predict morbidity of newly received calves in a commercial feedlot. Can. J. Anim. Sci. 81 (3), 315–320 (13 ref, Update Code: 20020000).
Rajkondawar, P.G., Tasch, U., Lefcourt, A.M., Erez, B., Dyer, R.M., Varner, M.A., 2002. A system for identifying lameness in dairy cattle. Appl. Eng. Agricult. 18 (1), 87–96 (19 ref, Update Code: 20020000).
Rajkondawar, P.G., Liu, M., Dyer, R.M., Neerchal, N.K., Tasch, U., Lefcourt, A.M., Erez, B., Varner, M.A., 2006. Comparison of models to identify lame cows based on gait and lesion scores, and limb movement variables. J. Dairy Sci. 89 (11), 4267–4275. 11.
Rasmussen, M.D., 2002. Defining acceptable milk quality at time of milking. In: McLean, J., Sinclair, M., West, B. (Eds.), First North American Conference on robotic milking, Wageningen, 20–22 March, 2002. Danish Institute of Agricultural Sciences, Foulum, DK-8830 Tjele, Denmark., Wageningen Pers. ISBN 90-76998-124. Source: First North American Conference on robotic milking, Toronto, Canada, 20-22 March, 2002; 2002.:IV-9-IV-16.; Update Code: 20020000.
Rasmussen, M.D., 2005. Visual scoring of clots in foremilk. J. Dairy Res. 72 (4), 406–414 (20 ref, Update Code: 20050000).
Rasmussen, M.D., Bjerring, M., 2005. Detection of clinical mastitis in automatic milking systems. ICAR Tech. Ser. (10), 117–121 (26–28 April, 2005 Update Code: 20100000).

Rutten, C.J., Velthuis, A.G.J., Steeneveld, W., Hogeveen, H., 2013. Invited review: sensors to support health management on dairy farms. J. Dairy Sci. 96 (4), 1928–1952 (Update Code: 20130417).

Rutten, C.J., Steeneveld, W., Inchaisri, C., Hogeveen, H., 2014. An ex ante analysis on the use of activity meters for automated estrus detection: to invest or not to invest? J. Dairy Sci. 97 (11), 6869–6887 (Update Code: 20141126).

Sherlock, R., Hogeveen, H., Mein, G., Rasmussen, M., 2008. Performance evaluation of systems for automated monitoring of udder health: analytical issues and guidelines. In: Lam, T.J.G.M. (Ed.), Mastitis Control: From Science to Practice. Proceedings of International Conference, Wageningen, 30 September–2 October 2008 2008. SmartWork Systems Ltd, Christchurch, New Zealand (ras@smartwork.co.nz, Mastitis control: from science to practice. Proceedings of International Conference, The Hague, Netherlands; Wageningen Academic Publishers. ISBN 978-90-8686-085-2. September - 2 October 2008; 2008.:275-282; Update Code: 20090000).

Sorensen, C.G., Fountas, S., Nash, E., Pesonen, L., Bochtis, D., Pedersen, S.M., Basso, B., Blackmore, S.B., 2010. Conceptual model of a future farm management information system. Comp. Electron. Agricult. 72 (1), 37–47 (Update Code: 20100000).

Steeneveld, W., Hogeveen, H., Barkema, H.W., van den Broek, J., Huirne, R.B.M., 2008. The influence of cow factors on the incidence of clinical mastitis in dairy cows. J. Dairy Sci. 91 (4), 1391–1402 (Update Code: 20080000).

Steeneveld, W., van der Gaag, L.C., Barkema, H.W., Hogeveen, H., 2009. Providing probability distributions for the causal pathogen of clinical mastitis using naive bayesian networks. J. Dairy Sci. 92 (6), 2598–2609 (ID: 58; Update Code: 20090000).

Steeneveld, W., van der Gaag, L.C., Ouweltjes, W., Mollenhorst, H., Hogeveen, H., 2010a. Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems. J. Dairy Sci. 93 (6), 2559–2568 (Update Code: 20100000).

Steeneveld, W., van der Gaag, L.C., Barkema, H.W., Hogeveen, H., 2010b. Simplify the interpretation of alert lists for clinical mastitis in automatic milking systems. Comp. Electron. Agricult. 71 (1), 50–56 (ID: 44; Update Code: 20100000).

Tanida, H., Koba, Y., Rushen, J., Passile, A.M., 2011. Use of three-dimensional acceleration sensing to assess dairy cow gait and the effects of hoof trimming. Anim. Sci. J. 82 (6), 792–800 (Update Code: 20120000).

Tello, A.S., Lokhorst, C., van Hertem, T., Halachmi, I., Maltz, E., Voros, A., Romanini, C. E.B., Viazzi, S., Bahr, C., Koerkamp, P.W.G.G., Berckmans, D., 2011. Selection of a golden standard for visual-based automatic lameness detection for dairy cows. In: Kofer, J., Schobesberger, H. (Eds.), Animal Hygiene and Sustainable Livestock Production. Proceedings of the XVth International Congress of the International Society for Animal Hygiene, Brno, 3–7 July 2011. Wageningen UR Livestock Research, Lelystad, Netherlands., Tribun EU. ISBN 978-80-263-0008-3. Volume 1; Source: Animal hygiene and sustainable livestock production. Proceedings of the XVth International Congress of the International Society for Animal Hygiene, Vienna, Austria, 3–7 July 2011, Volume 1; 2011.:325-327.; Update Code: 20110000.

Turner, S.P., Sinclair, A.G., Edwards, S.A., 2000. The interaction of liveweight and the degree of competition on drinking behaviour in growing pigs at different group sizes. Appl. Anim. Behav. Sci. 67 (4), 321–334 (Update Code: 20000000).

van Hertem, T., Maltz, E., Antler, A., Romanini, C.E.B., Viazzi, S., Bahr, C., SchlageterTello, A., Lokhorst, C., Berckmans, D., Halachmi, I., 2013. Lameness detection based on multivariate continuous sensing of milk yield, rumination, and neck activity. J. Dairy Sci. 96 (7), 4286–4298. ID: 89; Update Code: 20130717.

van Hertem, T., Viazzi, S., Steensels, M., Maltz, E., Antler, A., Alchanatis, V., SchlageterTello, A.A., Lokhorst, K., Romanini, E.C.B., Bahr, C., Berckmans, D., Halachmi, I., 2014. Automatic lameness detection based on consecutive 3d-video recordings. Biosyst. Eng. (ID: 87; Update Code: 20140326).

Viazzi, S., Bahr, C., Schlageter-Tello, A., Van Hertem, T., Romanini, C.E.B., Pluk, A., Halachmi, I., Lokhorst, C., Berckmans, D., 2013. Analysis of individual classification of lameness using automatic measurement of back posture in dairy cattle. J. Dairy Sci. 96 (1), 257–266. 1.

White, R.P., Schofield, C.P., Green, D.M., Parsons, D.J., Whittemore, C.T., 2004. The effectiveness of a visual image analysis (via) system for monitoring the performance of growing/finishing pigs. Anim. Sci. 78 (3), 409–418 (Update Code: 20040000).

Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Elsevier.

Wood, S., Lin, Y., Knowles, T.G., Main, D.C.J., 2015. Infrared thermometry for lesion monitoring in cattle lameness. Veter. Rec. 176 (12), 308 (Update Code: 20150429).

XiangYu, Song, Leroy, T., Vranken, E., Maertens, W., Sonck, B., Berckmans, D., 2008. Automatic detection of lameness in dairy cattle - vision-based trackway analysis in cow's locomotion (special issue: Smart sensors in precision livestock farming). Comp. Electron. Agricult. 64 (1), 39–44 (Update Code: 20080000).

6

PAPER II

Submitted for publishing in *Computers and Electronics in Agriculture*: Dominiak et al. (2017)

SPATIAL MODELING OF PIGS' DRINKING PATTERNS AS AN ALARM REDUCING METHOD I. DEVELOPING A MULTIVARIATE DYNAMIC LINEAR MODEL

K. N. Dominiak, L.J. Pedersen and A. R. Kristensen

Abstract: The overall objective of both the present and a following paper is to investigate spatial modeling of pigs' water consumption as an alarm reducing strategy for a future detection system in commercial pig production. In the present paper, the initial step is taken, and a spatial model is developed. For that purpose, the water consumption from multiple pens in multiple sections are monitored simultaneously by flow meters in both a commercial herd of finisher pigs (30-110 kg) and a research facility herd of weaner pigs (7-30 kg). The diurnal drinking patterns are modeled by a multivariate *dynamic linear model* (DLM), which is superpositioned by four sub-models describing three harmonic waves and a growth trend. Seven model versions reflect a variety of correlation structures between the monitored drinking patterns. The model versions were trained on learning data of the two herds, and run on separate test data sets from the herds. Their ability to fit the test data is measured as *mean square error* (MSE). Results for the finisher herd indicate correlation in data from pens within the same section (MSE = 13.850). For the weaner herd, results indicate an inverse relation between the degree of correlation and the model fit. Thus, the best fit (MSE = 1.446) is found for the model version expressing least correlation in data from pens across the herd. However, the estimated variance components indicate overfitting of the learning data, and the model fit may therefore not express the actual correlation. The present paper is the first part of two in the development of a spatial detection system. The application of the model to test data, and the evaluation of detection performance, is described in a subsequent article.

6.1 INTRODUCTION

The everyday focus in livestock production is to ensure a profitable production without compromising animal welfare. Over the years, livestock production has been subjected to an increasing industrialization, which has lead to larger, centralized production units with less time available for attending the individual animal (Berckmans, 2014; Sorensen et al., 2010).

Sensor-based monitoring and early warning systems can aid the daily manager to identify individual animals, or groups of animals, which need primary attention. Ideally the system can generate a warning timely enough for the manager to decide for the right intervention and either prevent any welfare reducing condition from occurring, or at least reduce its consequences (Kristensen et al., 2010). Such early warning systems, or detection models, for livestock production have been developed for the past twenty years (Dominiak and Kristensen, 2017), and they often aim to detect very specific conditions in individual animals, as for instance Clinical Mastitis (CM) in cows (Cavero et al., 2006, 2007; Huybrechts et al., 2014; Kamphuis et al., 2010; Mol et al., 1997, 1999), lameness (Garcia et al., 2014; Hertem et al., 2013, 2014; Kamphuis et al., 2013; Kramer et al., 2009; Maertens et al., 2011; Pastell and Kujala, 2007; Viazzi et al., 2013) and oestrus (Bressers et al., 1995; Cornou and Lundbye-Christensen, 2008; Freson et al., 1998; Maatje et al., 1997; Mol and Ouweltjes, 2001; Mol et al., 1997; Ostersen et al., 2010). However, the prevalence of animals with such specific conditions is usually low relative to the amount of animals not having them, and the consequence of this is that the warning systems generate too many false alarms (Dominiak and Kristensen, 2017; Hogeveen et al., 2010; Rasmussen, 2002).

For bio-security reasons, modern Danish pig production units for growing pigs are run very disciplined and systematically with a clear spatial separation between pigs of different age groups in closed sections (Cameron, 2000; Danish Agriculture and Food Council, 2010). This separation restrain most diseases from spreading between sections in a herd and, to a certain extent, between pens in a section (Cameron, 2000; Pedersen, 2012; Vils, 2013).

From a modeling perspective, such a construction of the production unit makes it well suited for the development of a spatial model. Hence, the herd can be modeled as a system consisting of one large unit (the whole herd), which consists of a number of identical subunits (sections), with each subunit consisting of a number of identical sub-subunits (pens). Such a spatial detection model aims to identify specific high-risk areas within the herd, rather than target individual animals. Area-specific alarms enables the manager to include any specific knowledge of the animals in the targeted areas, and hereby choose the best suited intervention under the given circumstances.

The parameter used in the model must contain relevant information on all animals across the herd in order to reflect the entire modeled system. Madsen et al. (2005) modeled the drinking pattern of a whole section of weaner pigs, and found that changes in the pattern contained information on the general wellbeing of the pigs as well as predictive value for detecting outbreaks of diarrhea. Later Andersen et al. (2016) showed that changes in drinking patterns could indicate stress caused by a variety of factors like stocking density and amount of rooting material supplied. These studies indicate a high level of information in water data, and this is supported in a recent study, where Jensen et al. (2017) found unexpected changes in the pigs' water consumption to be the one single parameter containing most information in the prediction of outbreaks of either diarrhea or fouling in a pen with finisher pigs.

Previous modeling of water data from growing pigs has been done on individual pens (Andersen et al., 2016; Jensen et al., 2017; Kashiha et al., 2013) or on the total water consumption in a section (Madsen et al., 2005). By modeling pens or sections separately, each modeled unit is considered isolated from other parts of the herd, whereas an incorporation of a herd-specific correlation between pens in the same section and sections in the same herd, could reflect interaction across the herd.

The objective of this paper is to present a spatial approach for modeling the drinking pattern of growing pigs throughout the entire growing period using a multivariate dynamic linear model. It is our hypothesis that pens and sections in a herd of growing pigs are correlated, and that this correlation can be modeled using model parameters defined at different spatial levels.

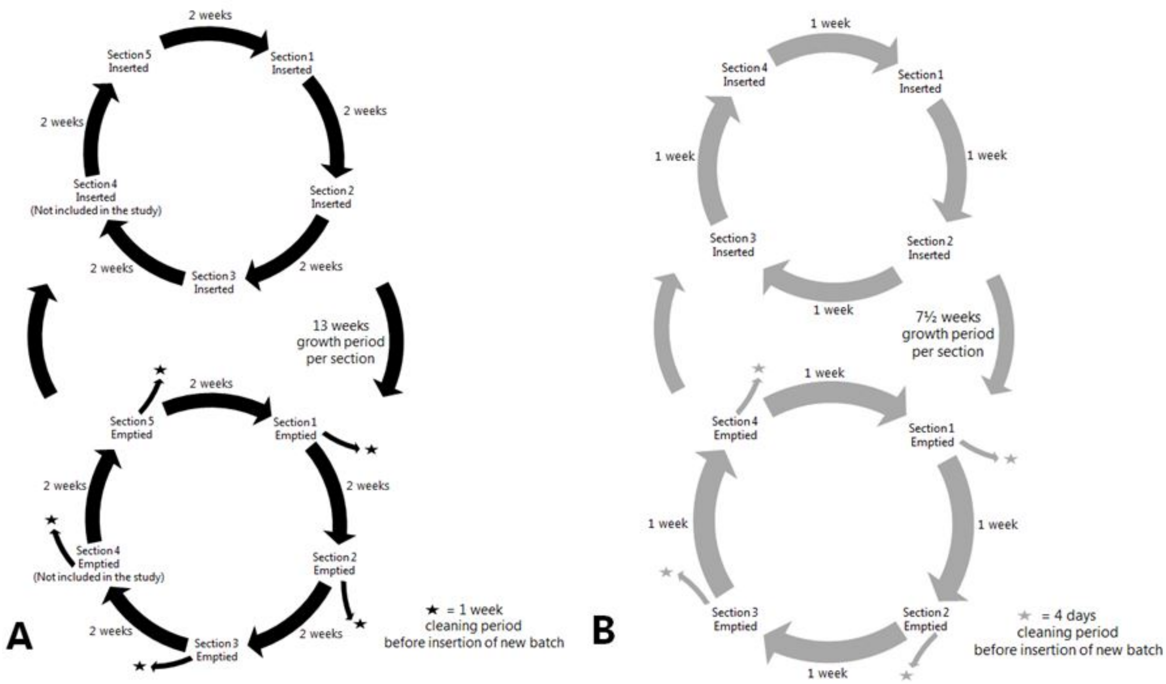


Figure 6.1: Production cycle for Herd A (A) and Herd B (B).

6.2 HERDS, SENSORS AND DATA

6.2.1 *Herd description*

For this study, water consumption data was obtained from two different herds. Herd A is a Danish commercial finisher herd, and Herd B is an experimental weaner herd, "Grønhøj", owned by the Danish Pig Research centre.

The general routines in Danish weaner and finisher production are structured so that the time of insertions of pigs in the farm, and the length of the growth period run in a cycle (Figure 6.1). Such a production cycle is a part of a larger production plan coordinated with the suppliers of the incoming pigs and the abattoir, when regarding finishers, or buyers, when regarding weaners. All pigs in one section are inserted at the same day, and they are all of same age relative to weaning date. When a section is emptied, it is cleaned and dried out for bio-security reasons before a new batch of pigs are inserted. For Herd A one growth period (30-110 kg) is approximately 14 weeks including one week of cleaning (Figure 6.1 (A)), and for Herd B one growth period (7-30 kg) is 8 weeks including four days of cleaning (Figure 6.1 (B)).

Herd A produces 10.000 cross-bred finisher pigs per year, and the herd has five identical sections, of which four are included in this study (Figure 6.2 (A)). Each section consists of 28 pens, and two neighbouring pens share the same water pipe, which supplies one drinking nipple in each of the pens (Figure 6.3). Approximately 486 pigs are inserted in a section with 18 pigs in each pen, and they are fed with liquid feed three times a day (Krogsdahl, 2014b). From 60 kg bodyweight the pigs are fed

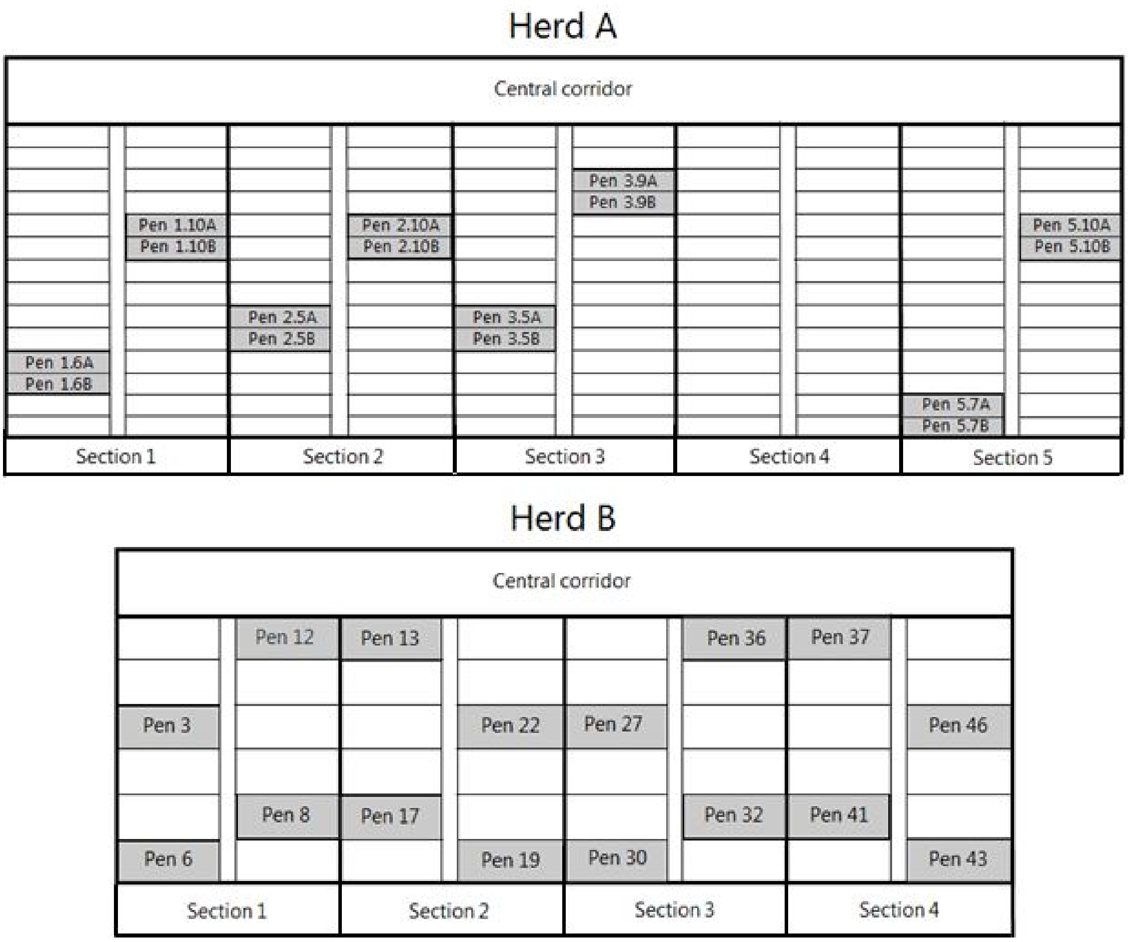


Figure 6.2: Structure of Herd A and Herd B. Grey double pens in Herd A and grey pens in Herd B were equipped with sensors and included in the study

restrictively as it is common practice with finisher pigs in order to increase the lean meat percentage (Vils, 2012).

Herd B consists of four sections, each with 12 pens for weaner pigs (Figure 6.2 (B)). One water pipe supplies one drinking bowl per pen (Figure 6.3). 15 pigs are inserted in each pen, and the pigs are fed ad libitum with dry feed three times a day during the whole growth period (Krogsdahl, 2014a).

The main characteristics of the two herds are summarized in Table 6.1.

6.2.2 *Data*

Water data was obtained by photo-electric flow sensors (RS V8189 15mm Diameter Pipe) measuring water flow per millisecond as pulses proportional to the velocity of the water (Anonymous, 2000). The sensors were calibrated between batches, and the number of pulses entered a central data base once every 24 hours. For this study the number of pulses were converted to litres and aggregated per hour, yielding water use in litres per hour.

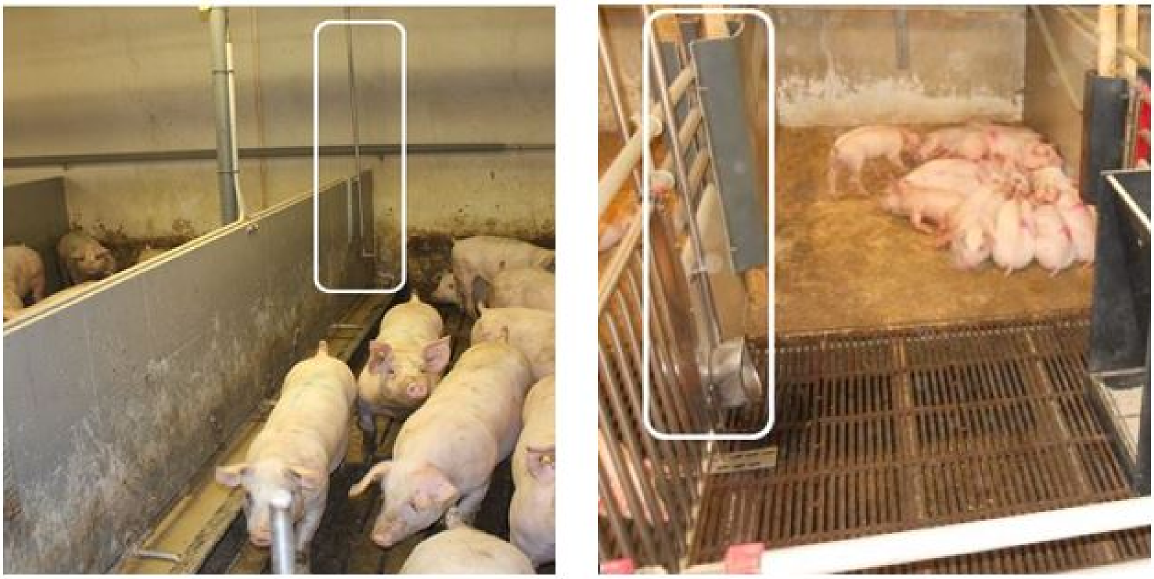


Figure 6.3: One water pipe supplying two neighbouring pens in Herd A (left) and a single pen in Herd B (right)

Table 6.1: Characteristics for the two herds in the study (14 for section K11)

Characteristic	Herd A	Herd B
Production type	Commercial	Research Farm
Animal group	Finishers (30-110 kg)	Weaners (7-30 kg)
Sections	4	4
Sensors total/ per section	8/2	16/4
Pigs per pen/ per sensor	18/36	15/15
Growth period (batch)	14 weeks	8 weeks
Batches per sensor	7	13 [1]
Learning data (hours)	9540	14657
Test data (hours)	4441	3025

[1] 14 for section 4.

In Herd A a total of eight sensors were installed with two sensors placed in each of four identical sections (Figure 6.2 (A)). All sensors were placed on water pipes supplying two neighbouring pens, and therefore each sensor monitored the joint water use of pigs in two pens. Both sections and pens were randomly chosen, and seven batches were monitored per section from May 2014 to March 2016. In this study a batch of pigs is defined as all pigs inserted in the same section at the same day.

The full data set for Herd A consists of eight time series, one per sensor, of length from the first observation in the herd to the last observation in the herd. In total 16309 hours. Every observation from each sensor is paired with the *insertion date* of the relevant batch of pigs at any given time. The data set for Herd A was divided into a learning data set, which consists of the first four batches (9540 hours, 68%) and a test data set, which consists of the two last batches (4441 hours, 32%) (Table 6.1). A total period of 2328 hours (one batch per pen) was left out between the two subsets to exclude the possibility of observations from the same pigs occurring in both data subsets.

In Herd B a total of sixteen sensors were installed with four sensors in each of four identical sections (Figure 6.2 (B)). Each sensor monitored the water use of one individual pen. The sections included in this study were assigned by the research centre, whereas the pens within each section were randomly chosen. 13 batches (Sections 1, 2, and 3) and 14 batches (Section 4) were included, and data was collected from October 2014 to December 2016.

The full data set for Herd B consists of sixteen time series, one per sensor. The monitoring period begins with the first global observation and ends with the final global observation. In total 18755 hours. As in the data set for Herd A, every observation is paired with the insertion date of the relevant batch. The data set for Herd B was divided into a learning data set of the first ten batches (14657 hours, 83%) and a test data set of the two last batches (3025 hours, 17 %) (Table 6.1). A period of 1073 hours (one batch) was left out between the two data subsets to ensure no observations from the same pigs would occur in both subsets.

During cleaning periods between batches, no sensor observations were made. Such periods were considered planned periods of missing data as opposite to any occasional missing observations or sensor outages during the growth periods.

As only actual water flow is measured, it is not possible to distinguish periods with no water consumption from (short) sensor outages. Since water consumption is typically very low during the night, it was decided to interpret missing observations of a duration of less than 5 hours between 10:00 PM and 4:00 AM as zero observations. All other missing observations are considered as sensor outages.

6.3 MODEL DESCRIPTION

In this section the structure of the developed model is described. The model is developed as a general tool, which in theory can be applied to any herd with either weaner pigs or finisher pigs.

6.3.1 *General Dynamic Linear Model*

The water consumption over time is modeled simultaneously for all sensors in the herd. The observation vector $Y_t = (Y_{1t}, \dots, Y_{nt})'$ is the water consumed within the last hour at time t for each of the n sensors. It is modeled by the matrix quadruple $\mathbf{F}_t$, $\mathbf{G}_t$, $\mathbf{V}_t$, and $\mathbf{W}_t$, where, following the description by West and Harrison (1999):

- $\mathbf{F}_t$ is a known $(n \times r)$ design matrix;
- $\mathbf{G}_t$ is a known $(n \times n)$ system matrix;
- $\mathbf{V}_t$ is a known $(r \times r)$ observation variance-covariance matrix;
- $\mathbf{W}_t$ is a known $(n \times n)$ system variance-covariance matrix.

The four matrices, $\mathbf{F}_t$, $\mathbf{G}_t$, $\mathbf{V}_t$, and $\mathbf{W}_t$, define the way $\mathbf{Y}_t$ relates to an underlying parameter vector θ_t at time t, and how the system evolves over time in the two equations:

OBSERVATION EQUATION

$$Y_t = \mathbf{F}_t'\theta_t + \nu_t, \quad \nu_t \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_t), \tag{10}$$

and

SYSTEM EQUATION

$$\theta_t = \mathbf{G}_t'\theta_{t-1} + \omega_t, \quad \omega_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t). \tag{11}$$

The aim of the DLM is to estimate the parameter vectors $\theta_1, \dots, \theta_t$ from the observations $Y_1, \dots, Y_t$ by sequential use of the Kalman filter. Let D_0 denote the initial information before any observations are made so that $(\theta_0|D_0) \sim \mathcal{N}(m_0, C_0)$. Furthermore, let $D_{t-1} = D_0 \cup \{Y_1, \dots, Y_{t-1}\}$ denote all available information before time t so that $(\theta_{t-1}|D_{t-1}) \sim \mathcal{N}(m_{t-1}, C_{t-1})$. When a new observation Y_t becomes available, the Kalman filter will update the conditional distribution from $\mathcal{N}(m_{t-1}, C_{n-1})$ to $\mathcal{N}(m_t, C_n)$ as described by West and Harrison (ibid.).

6.3.2 *Model construction*

When looking at Figures (6.4) and (6.5) it can be seen that the water consumption of growing pigs has a clear diurnal pattern. Furthermore, Figure (6.6) illustrates how the underlying level of water consumed per day increases over time, implying that the underlying level of daily water consumption increases as the pigs grow.

Madsen et al. (2005) found that the drinking pattern of a whole section of 405 weaner pigs could be described in a DLM composed of four smaller DLMs, describing three harmonic waves of lengths

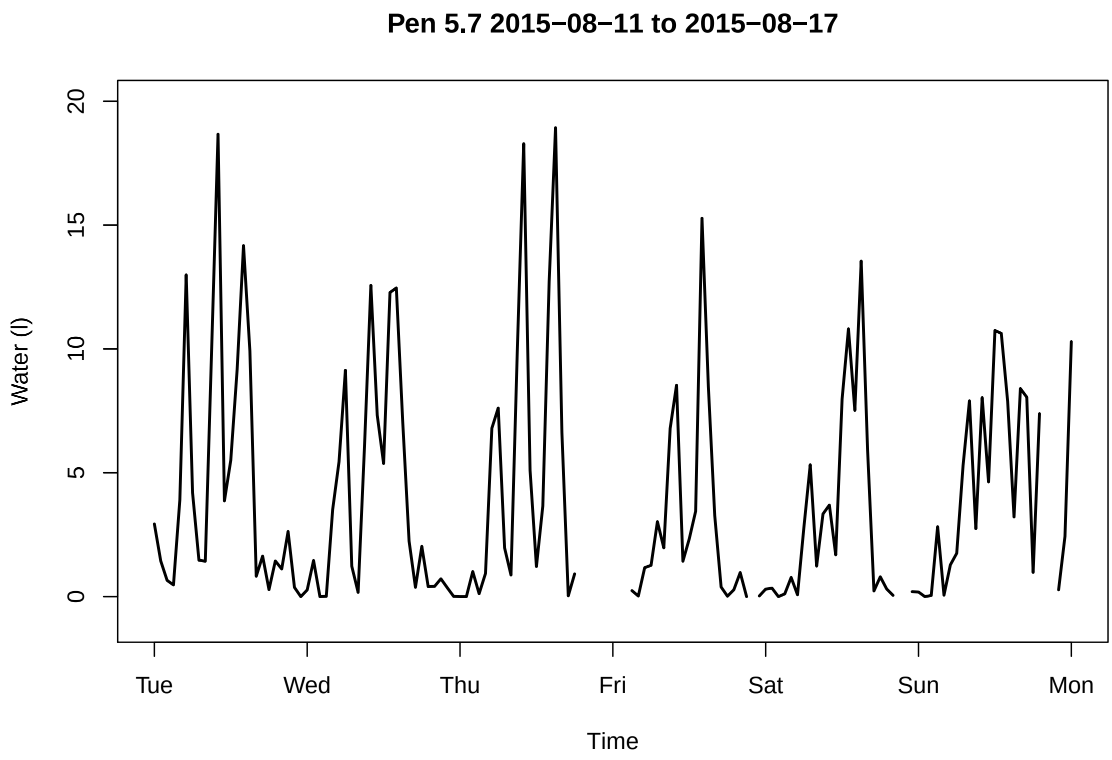


Figure 6.4: Diurnal drinking pattern, finishers Herd A.

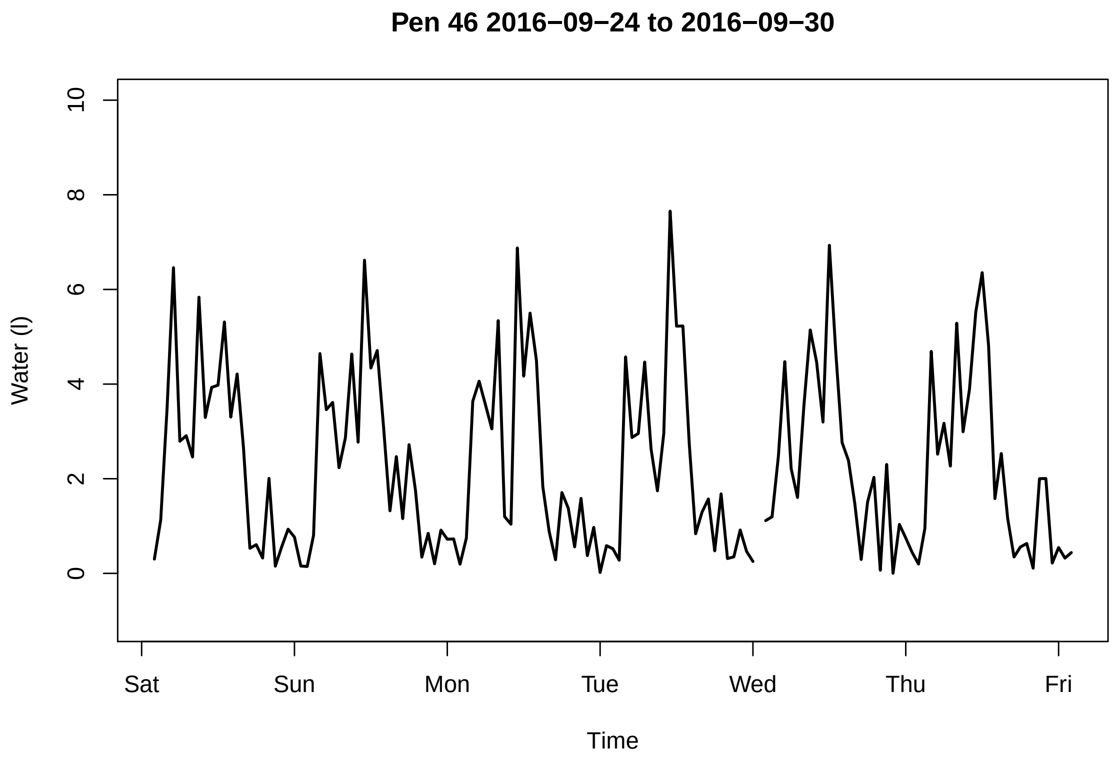


Figure 6.5: Diurnal drinking pattern, weaners Herd B.

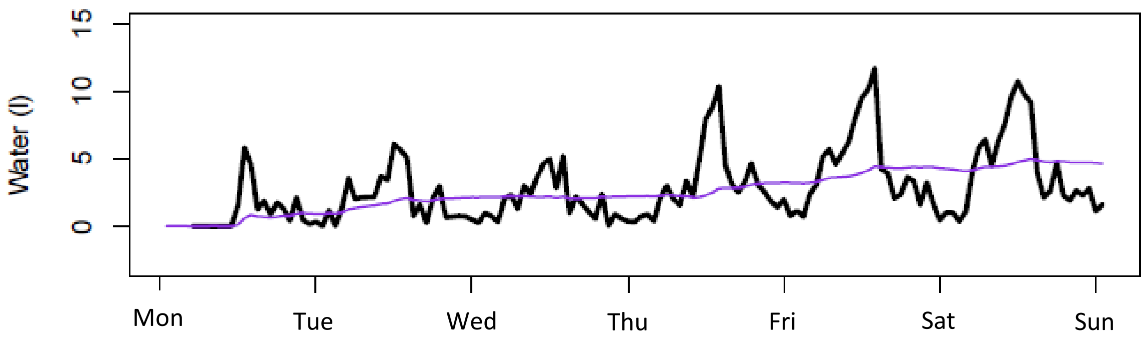


Figure 6.6: Drinking pattern of one week (black line) where the underlying level increases over time (purple line).

24h, 12h, and 8h, and a growth trend. The same four sub models describe the diurnal drinking pattern of a pen of weaners or finishers very good as well, as illustrated in Figure (6.7), and the development of the full multivariate model will be described in the following subsections.

Cyclic models

The diurnal drinking pattern is modeled by three cyclic models, each describing a harmonic wave. Harmonic waves can be expressed in a DLM using trigonometric functions in the *Fourier form representation of seasonality* (Madsen et al., 2005; West and Harrison, 1999), where each wave takes up two parameters, representing the phase and amplitude of the cosine waveform. According to West and Harrison (1999), the harmonic waves can be described with the design matrix $\mathbf{F}_t^h$ and system matrix $\mathbf{G}_t^h$ defined as:

$$\mathbf{F}_t^h = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{G}_t^h = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}. \tag{12}$$

with $\omega = 2\pi/24$ yielding a wave with a period of 24 (M_{H1}), $\omega = 2\pi/12$ a wave with a period of 12 (M_{H2}), and $\omega = 2\pi/8$ a wave with a period of 8 (M_{H3}).

Linear growth model

The underlying level of water consumption can be described by a linear function, and the increase over time is included by combining the linear function with a growing trend in a *linear growth model*, modeling the increase from time $t-1$ to t.

The general description of a dynamic linear growth model, as based on West and Harrison (ibid.), is characterized by the following design and system matrices:

$$\mathbf{F}_t^l = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{G}_t^l = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{13}$$

The parameter vector θ_t consists of a level parameter θ_{1t} and a growth parameter θ_{2t}. Thus, the expected level at time, θ_{1t}, will be the sum of the level at time $t-1$ and the growth parameter.

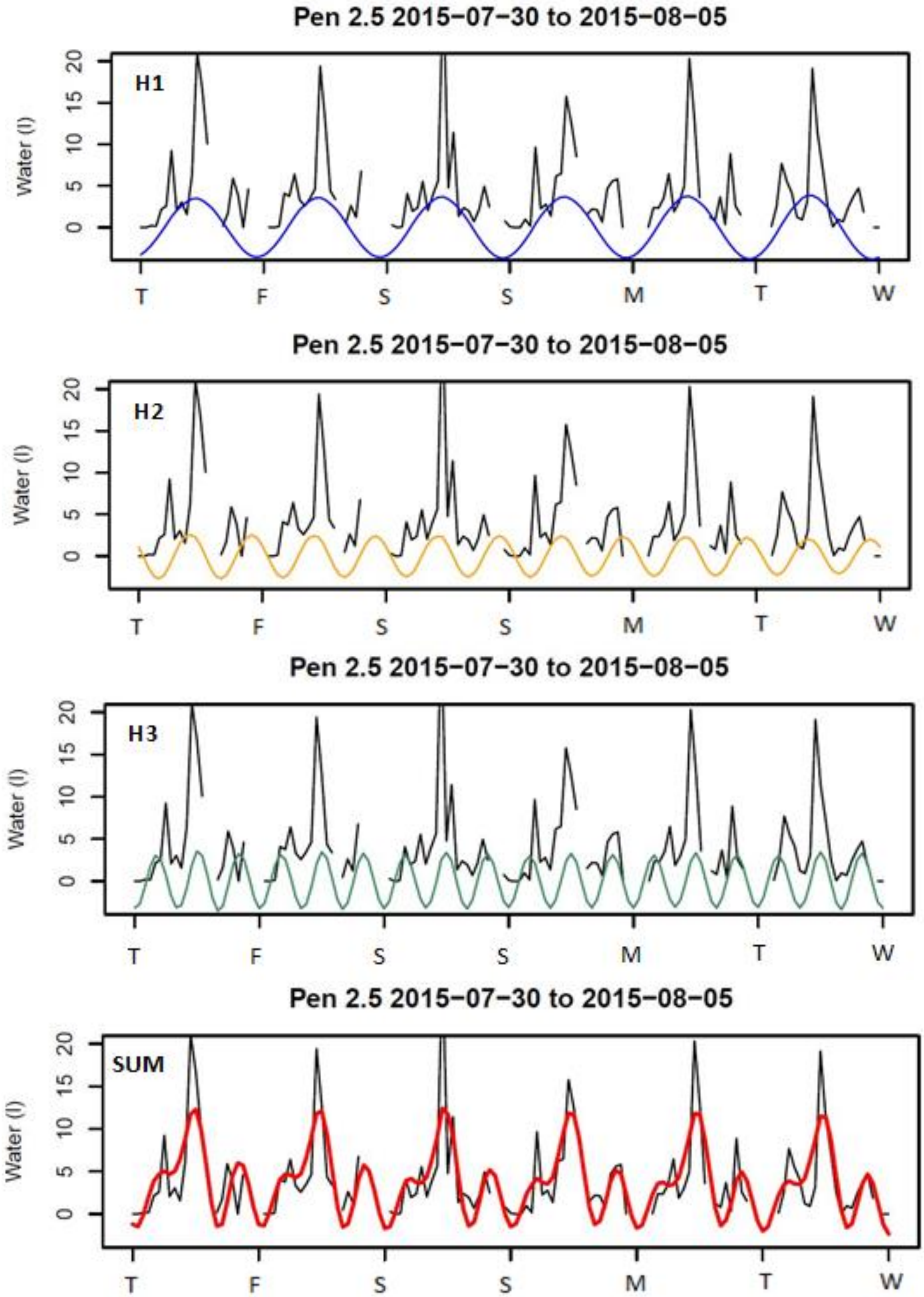


Figure 6.7: The diurnal drinking pattern (black line) is shown together with the three harmonic waves; 24 h (H1), 12 h (H2), and 8 h (H3). The sum of the three harmonic waves and the underlying level (which is not depicted) is shown in (SUM).

6.3.3 *Full model - univariate*

For a single sensor, the univariate model consisting of four sub models; one linear growth model M_{LG} and the three cyclic models M_{H1}, M_{H2} and M_{H3}, is characterized by the design matrix

$$\mathbf{F}_t^u = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}' \tag{14}$$

and the system matrix

$$\mathbf{G}_t^u = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos(\omega) & \sin(\omega) & 0 & 0 & 0 & 0 \\ 0 & 0 & -\sin(\omega) & \cos(\omega) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos(2\omega) & \sin(2\omega) & 0 & 0 \\ 0 & 0 & 0 & 0 & -\sin(2\omega) & \cos(2\omega) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cos(3\omega) & \sin(3\omega) \\ 0 & 0 & 0 & 0 & 0 & 0 & -\sin(3\omega) & \cos(3\omega) \end{pmatrix} \tag{15}$$

where $\omega = 2\pi/24$.

The observation variance-covariance matrix reduces in the univariate case to a scalar, V_t^u, whereas the size of the system variance-covariance matrix, $\mathbf{W}_t^u$, is of size 8×8. The parameter vector θ_t has eight elements; one for level, one for growth, and two for each of the three harmonics.

Full model - multivariate

The simplest possible multivariate model for n sensors would be to define a system matrix $\mathbf{G}_t$ of size $8n \times 8n$ as a block diagonal matrix where each of the n blocks along the diagonal is identical to G_t^u from Eq. (15). Similarly, the observation variance-covariance matrix would be a diagonal matrix having all diagonal elements equal to V_t^u. The system variance-covariance matrix would be a block diagonal matrix where each block along the diagonal would be equal to $\mathbf{W}_t^u$. Finally, the design matrix $\mathbf{F}_t$ would be a $8n \times n$ matrix with n blocks each equal to $\mathbf{F}_t^u$. The underlying assumption behind such a model would be that the observations from the n sensors were completely independent.

A multivariate model as described would, however, not add anything to a scenario with n univariate models running separately in parallel. Therefore, the model has to be modified to allow for interactions between sensors. This can, basically, be achieved by direct modeling of the interactions in the design and system matrices and/or by estimating full variance-covariance matrices $\mathbf{V}_t$ and $\mathbf{W}_t$ allowing for correlations between sensors (as opposed to block diagonal matrices).

In this study both approaches will be used. The interactions between the elements of the parameter vector will be directly modeled in the design and system matrices, and the interactions between the observation errors will be modeled by a full variance-covariance matrix.

Modeling interactions in the design and system matrices

When modeling the spatial structure of the two herds included in this study, three spatial levels, Pen, Section and Herd, are defined as follows:

- *Pen Level* describes individual sensors, each monitoring either the joined water consumption of pigs in two neighbouring pens (Herd A), or in a single pen (Herd B).
- *Section Level* describes each of the physical sections of a herd as one individual unit including all sensors within the section.
- *Herd Level* describes the physical building in which the sections are placed, and includes all sensors in all pens and sections.

In order to describe any herd-specific interactions, the four sub models can be defined individually at either of the three spatial levels with the following properties:

- Defined at *Pen Level* the sub model evolves differently in each pen over time. One block is formed per pen in $\mathbf{G}_t$.
- Defined at *Section Level* the sub model evolves identically in all pens within the section over time, expressing interactions between pens in the same section. One block is formed per section in $\mathbf{G}_t$.
- Defined at *Herd Level* the sub model evolves identically in all pens in the herd over time, expressing interactions between all pens in the herd. One block including the whole herd is formed in $\mathbf{G}_t$.

Let M_{iLv} denote a sub model, i, where $i \in \{LG, H1, H2, H3\}$ indicates either the linear growth (LG), or one of the three cyclic models ($H1, H2, H3$). Furthermore, let a sub model be defined at level Lv, where $Lv \in \{h, s, p\}$ for either herd (h), section (s), or pen (p). Finally, n_s and n_p denote the number of sections in the herd and the number of pens in the herd, respectively.

As an example of the direct modeling of interactions in the design and system matrices; take a herd with two sections ($n_s = 2$) each with two sensors (so that $n_p = 2n_s = 4$), and the following definitions of interactions for each sub model:

- M_{LGs} is defined at *Section Level*
- M_{H1h} is defined at *Herd Level*
- M_{H2s} is defined at *Section Level*
- M_{H3p} is defined at *Pen Level*

The design matrix $\mathbf{F}_t^{ex}$ will then have dimensions 18×4 with the following structure:

$$\mathbf{F}_t^{ex} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}' \tag{16}$$

And the corresponding system matrix $\mathbf{G}_t^{ex}$ will have the dimensions 18×18, consisting of nine blocks in the following block diagonal structure:

$$\mathbf{G}_t^{ex} = \begin{pmatrix} G_{LGs}^{s1} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & G_{LGs}^{s2} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & G_{H1h}^{h} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & G_{H2s}^{s1} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & G_{H2s}^{s2} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & G_{H3p}^{s1p1} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & G_{H3p}^{s1p2} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & G_{H3p}^{s2p1} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & G_{H3p}^{s2p2} \end{pmatrix} \tag{17}$$

where $\underline{0}$ denotes a 2×2 sub matrix only consisting of zeros.

Modeling interactions in the Observation variance-covariance matrix, $\mathbf{V}_t$

The observational variances, **Vt**, express the natural completely random variation of water consumption as well as any uncertainty in the measurements, or observations, of the water data. The full variance-covariance matrix for the observation error ν_t will be defined from three separate variance components corresponding to herd level, section level and pen level, respectively. This corresponds to an error structure as follows for the jth sensor (placed in Pen P_j, located in Section S_j of Herd H_j):

$$\nu_{jt} = \nu_{H_j t} + \nu_{S_j t} + \nu_{P_j t}, \tag{18}$$

where

- $\nu_{H_j t} \sim \mathcal{N}(0, \sigma_H^2)$ is an error term which is common for the entire herd;
- $\nu_{S_j t} \sim \mathcal{N}(0, \sigma_S^2)$ is an error term which is common for all pens in a section;
- $\nu_{S_j t} \sim \mathcal{N}(0, \sigma_P^2)$ is an error term which is specific for a pen.

As an example, consider again a herd with two sections, each holding two pens with a sensor in each of them. The distribution of the observation error then becomes

$$\nu_{jt}^{ex} \sim \mathcal{N}\left(\left(\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \end{array}\right), \left(\begin{array}{cccc} \sigma_H^2+\sigma_S^2+\sigma_P^2 & \sigma_H^2+\sigma_S^2 & \sigma_H^2 & \sigma_H^2 \\ \sigma_H^2+\sigma_S^2 & \sigma_H^2+\sigma_S^2+\sigma_P^2 & \sigma_H^2 & \sigma_H^2 \\ \sigma_H^2 & \sigma_H^2 & \sigma_H^2+\sigma_S^2+\sigma_P^2 & \sigma_H^2+\sigma_S^2 \\ \sigma_H^2 & \sigma_H^2 & \sigma_H^2+\sigma_S^2 & \sigma_H^2+\sigma_P^2+\sigma_S^2 \end{array}\right)\right) \tag{19}$$

All three variance components of **Vt** are assumed constant over time for all batches, but different between herds. Leaking drinking bowls or drinking nipples often occur for a shorter period of time in one or a few pens, and this is likely to affect the pen level variance, which is also assumed for any inaccuracies of the flow meters. Both mixing of liquid feed (in Herd A) and washing of sections between batches occur with the same frequency for all batches, affecting all pigs in the herd, and therefore assumably the herd level variance. Also a possible influence of weather conditions is assumed to be expressed at herd level.

Modeling interactions in the System variance-covariance matrix, $\mathbf{W}_t$

The system variance, $\mathbf{W}_t$, expresses any uncertainty about the changes of the state vector from time $t-1$ to t, and hereby determines the stability of the system over time. As described by West and Harrison (1999) the system variance $\mathbf{W}_t$ can be expressed as a fixed proportion of the estimated variance $\mathbf{C}_t$ of θ_t given all observations until (and including) time t by a *discount factor*, δ, which by definition, satisfies the condition $0 < \delta \leqslant 1$ (ibid.). Given δ and $\mathbf{C}_0$ the whole series $\mathbf{W}_t$ can be identified as follows for each time step t:

$$\mathbf{W}_t = \frac{1-\delta}{\delta}\mathbf{P}_t, \tag{20}$$

where

$$\mathbf{P}_t = \mathbf{G_t C_{t-1} G_t'} = V[\mathbf{G_t}\theta_{t-1}|D_{t-1}], \tag{21}$$

However, one single discount factor is not recommended for a super positioned model (ibid.). Instead each sub model should be allowed to express different rates of change in stability over time through individual discount factors. In a diurnal pattern the harmonic characteristics are often more durable than the growth trend, which further emphasizes the need for a discount factor for each sub model (ibid.). We will therefore need several discount factors to express the system variance in the present model.

Based on West and Harrison (ibid.) the system variance of a super positioned DLM, can be defined using several discount groups as follows: Let M_i denote a sub model (i.e. a certain range of parameters of the parameter vector θ_i). Let $\gamma \geqslant 1$ denote the number of sub models and let n_i denote the number of parameters of M_i, so that $\sum_{i=1}^{\gamma} n_i = n$ is the dimension of the full super positioned DLM.

A block diagonal approach is then applied where, for instance, G_{it} denotes the ith block diagonal element of G_t. Thus, for sub model M_i, we have

$$\mathbf{P}_{it} = \mathbf{G}_{it}\mathbf{C}_{i,t-1}\mathbf{G}'_{it} = V[\mathbf{G}_{it}\theta_{i,t-1}|D_{t-1}], \qquad i = 1, \ldots, \gamma \tag{22}$$

and

$$\mathbf{W}_{it} = \frac{1-\delta_i}{\delta_i}\mathbf{P}_{it}, \qquad i = 1, \ldots, \gamma, \tag{23}$$

where δ_1,...,δ_γ are any set of discount factors, $(0 < \delta_i \leqslant 1; i = 1, \ldots, \gamma)$, with δ_i being the discount factor associated with the sub model in question.

As described in Section 6.3.3, the spatial level of a sub model determines the number of blocks in the system matrix $\mathbf{G}_t$, and hereby the number of blocks in the system variance matrix $\mathbf{W}_t$ if there is no correlation between the sensors and the levels. The system variances for all blocks from the same sub model belong to the same *discount group*, meaning, they are expressed through the same discount factor. Since, however, some sensors are placed within the same section, and all sensors are placed within the same herd, some interactions between the changes of the corresponding state variables in the state vector are assumed:

- *Pens* are assumed dependent within the same section.
- *Sections* are assumed dependent within the herd.
- *Pens* are assumed independent between sections except for the correlation expressed through the sections.
- The linear growth model and the cyclic models are assumed independent of each other.

As a consequence of the defined correlation structure, the block structure for the example of Section 6.3.3 with two sections each with two sensors will result in $\gamma = 5$ blocks where

$$G_{1t} = \begin{pmatrix} G^{s1}_{LGs} & \underline{0} \\ \underline{0} & G^{s2}_{LGs} \end{pmatrix}, \quad G_{2t} = G^{h}_{H1h}, \quad G_{3t} = \begin{pmatrix} G^{s1}_{H2s} & \underline{0} \\ \underline{0} & G^{s2}_{H2s} \end{pmatrix},$$

$$G_{4t} = \begin{pmatrix} G^{s1p1}_{H3p} & \underline{0} \\ \underline{0} & G^{s1p2}_{H3p} \end{pmatrix} \quad \text{and} \quad G_{5t} = \begin{pmatrix} G^{s2p1}_{H3p} & \underline{0} \\ \underline{0} & G^{s2p2}_{H3p} \end{pmatrix}.$$

Thus, G_{2t} is a 2×2 matrix, whereas G_{1t}, G_{3t}, G_{4t} and G_{5t} are all 4×4 matrices. The system variance matrix will have the same block structure with 5 diagonal blocks of the same dimensions (but without the zeros).

With the structure defined above it would be possible to use 5 different discount factors, but in general, the same discount factors are used for all harmonic models with thew same wave length. Thus, Blocks 4 and 5 will in the example have the same discount factor reducing the number of different discount factors to 4.

6.4 APPLYING THE DLM

6.4.1 *Defining a herd*

Throughout the development of the multivariate spatial DLM just described, it has been the aim to build a very general model in a way so it can be applied to a variety of herds, not regarding the number of pens and sections, levels of sub models, lengths of batches, sizes of pigs etc. In this section we will describe how the model is applied to the data, but also illustrate the flexibility of the general model by applying it to the data sets from both a finisher herd with eight sensors, Herd A, and a weaner herd with sixteen sensors, Herd B.

In the model, a herd is defined by the number of sensors monitoring water consumption. Input to the model is the number of sections with sensors, the number of sensors within the section, and a list of which pen and section each sensor is placed in. In order to distinguish batches from each other, the insertion date of all batches related to each sensor is also given as input to the model.

6.4.2 *Handling missing observations*

Both data sets are characterized by periods of missing observations, which can involve everything from one to all sensors in the herd, and last from one hour to a whole batch. Any period of missing observations is registered as NA observations in the data set, and is handled individually for each sensor by the model. Some NA observations are related to the cleaning period between two batches, where no observations are registered from any sensor in the empty section. These periods are regarded as *planned periods* of missing observations, and at the insertion of a new batch the model is reset as will be described in Section 6.4.3. Some NA observations are, however, *unplanned missing observations*. A period of unplanned NA observations can be caused by the pigs not drinking any water, or by sensor outages for a shorter or longer period. During a sequence of NA observations, the system keeps evolving and $\mathbf{C}_t$ increases.

6.4.3 *Resetting between batches*

All data from all batches monitored by one sensor represent one long time series. However, each new batch is likely to evolve differently over time. Therefore each time series must be divided into subseries with the length of the specific batch. If a new batch is inserted at time t_n the learned values for conditional mean, m_t, and variance-covariance matrix, $\mathbf{C}_t$, from the previous batch are discarded, and the values are reset to $m_{t_n} = m_0$ and $\mathbf{C}_{t_n} = \mathbf{C}_0$.

6.4.4 *Estimating variance components*

As described in Sections 6.3.3 and 6.3.3, the observation variances are modeled through full variance-covariance matrices, whereas the system variances are modeled as a fixed proportion of the posterior variances, $\mathbf{C}_t$, using discount factors.

Any difference between the predicted multivariate observation and the actual observation is contained in a vector of forecast errors e_t. If the pigs follow their normal drinking pattern and drink as much water as expected, the prediction of the next observation is close to perfect, and any forecast error, e_t, will be small. If, on the other hand, something is causing the pigs to drink more or less than expected, the forecast error will be larger. The *mean square error*, MSE defined as $\frac{1}{T}\sum_{t=1}^{T} e_t' e_t$, will be used to measure the predictive performance of each model for comparison of the different versions.

All observation variances and discount factors are estimated on learning data by numerical optimization using the Nelder-Mead algorithm implemented in the `optim` function in R (R Core Team, 2014). The criterion of optimality is minimization of MSE. Thus, the observation variances and discount factors, which minimize the MSE for the learning data, are the results of the estimation. After insertion of a batch and after a sequence of NA observation lasting more than five hours, the model parameters need to adjust to the observations before reliable forecasts can be produced. Therefore, forecast errors for the first 24 hours after insertion and after such an NA sequence are ignored in the evaluation of the MSE.

6.4.5 *Model scenarios*

Now the general structure of the spatial DLM is described, and seven different versions of the full DLM will be applied to each of the two data sets in order to identify, which versions describe any spatial correlation in the drinking pattern of finishers and of weaners the best. Each model differs with regard to the defined levels of the cyclic models (see Table 6.2), and the two data sets are divided into learning sets and test sets, as described in Section 6.2.2. The observation variances, $\mathbf{Vt}$, and the discount factors, δ, of each model will be estimated on learning data, and the estimated values will be entered as input to the model, when run on test data.

As previously described, Danish pig producing units for growing pigs are generally run with a sectionized structure. Therefore pigs of different ages and sizes are located in different, and separate, sections of a herd. In order to reflect that, the linear growth sub model will be defined at section level for all seven versions in each herd, but the cyclic sub models of the DLM will be defined at different level combinations, which can be seen in Table 6.2

For both herds, and all model versions, the same four discount groups were defined; namely one for each of the four sub models.

Table 6.2: Level definitions for the seven models applied to data sets from Herd A and Herd B. The Linear Growth sub model is defined at Section level in all models, whereas different combinations of level definitions are made for the cyclic sub models. Notations: LG = Linear Growth model, H1 = Cyclic model of length 24, H2 = Cyclic model of length 12, H3 = Cyclic model of length 8. H = Herd level, S = Section level, P = Pen level

LG	H1 H2 H3	Interpretation
S	HHH	The full harmonic pattern evolves identically for all pens in the herd
S	HSS	H1 evolves identically for all pens, H2 and H3 evolve identically within each section but differently between sections
S	HSP	H1 evolves identically for all pens, H2 evolves identically within sections but differently between sections, H3 evolves differently in each pen
S	SSS	The full harmonic pattern evolves identically within each section but differently between sections
S	SSP	H1 and H2 evolve identically within sections but differently between sections, H3 evolves differently in each pen
S	SPP	H1 evolves identically within sections but differently between sections, H2 and H3 evolve differently in each pen
S	PPP	The full harmonic pattern evolves differently in each pen

Table 6.3: Estimated observation variances and discount factors for seven model structures for Herd A. Learning data: 9540 hours (68 %). Test data: 4441 hours (32%). Notations: LG = Linear Growth model, H1 = Cyclic model of length 24, H2 = Cyclic model of length 12, H3 = Cyclic model of length 8. H = Herd level, S = Section level, P = Pen level

Model Structure		Observation variances [1]			Discount factors [2]			
LG	H1 H2 H3	σ^2_H	σ^2_S	σ^2_P	δ_1	δ_2	δ_3	δ_4
S	HHH	5.84	146.72	13.34	0.993	0.989	0.994	0.990
S	HSS	9.29	176.09	0.12	0.993	0.989	0.993	0.992
S	HSP	0.60	90.39	172.45	0.993	0.989	0.994	0.994
S	SSS	0.10	104.41	0.10	0.992	0.994	0.992	0.9999
S	SSP	3.59	37.37	121.72	0.994	0.990	0.994	0.994
S	SPP	1.28	12.87	152.15	0.994	0.990	0.995	0.994
S	PPP	0.13	3.56*E-6	78.09	0.994	0.991	0.995	0.994

[1] σ^2_H: Observation variance, herd effect, σ^2_S: Observation variance, section effect, σ^2_H: Observation variance, pen effect
[2] δ_1: LG, δ_2: H1, δ_3: H2, δ_4: H3

6.5 RESULTS AND DISCUSSION

The estimated variance components and the predictive performance (evaluated as MSE) for each of the seven model versions is presented for both herds.

6.5.1 *Estimated variance components*

Tables 6.3 and 6.4 show the estimated variance components for each of the seven models applied to data from Herd A and Herd B. In the majority of the models, the observational variances, σ^2_H, σ^2_S and σ^2_P, are very high, especially σ^2_P for Herd B, but also σ^2_S and σ^2_P for Herd A. In general, high observational variances in a DLM can indicate that all variation in the data is expressed through the observation variances, assumably because of a very rigid and non-flexible system.

However, a rigid system would favour very high discount factors (i.e. δ very close to 1), leaving little room for any instability in the system variances (West and Harrison, 1999). This is not the case for the discount factors of this model, as it can be seen in Tables 6.3 and 6.4. Especially for Herd B, the estimated discount factors are fairly low, which results in a highly flexible model on all parameters, capable of adjusting very well to the learning data set (Witten and Frank, 2005).

The high flexibility, which characterize the models for both herds, can also be caused by too high a complexity of the models, that is an excess of parameters used to describe the data, hereby leading to an overfitting of data (Elith et al., 2008; Hawkins, 2004; Torgo, 2017; Witten and Frank, 2005). Overfitting often occurs in models with highly correlated parameters (Hawkins, 2004; Witten and Frank, 2005). The high observational variances we see in the DLM presented here, can be the result of a so-called bias-variance tradeoff, which is associated with overfitting (Torgo, 2017). The bias-

Table 6.4: Estimated observation variances and discount factors for seven model structures for Herd B. Learning data: 14657 hours (83 %). Test data: 3025 hours (17%). Notations: LG = Linear Growth model, H1 = Cyclic model of length 24, H2 = Cyclic model of length 12, H3 = Cyclic model of length 8. H = Herd level, S = Section level, P = Pen level

Model Structure		Observation variances [1]			Discount factors [2]			
LG	H1 H2 H3	σ^2_H	σ^2_S	σ^2_P	δ_1	δ_2	δ_3	δ_4
S	HHH	1.02	0.66	1486.23	0.973	0.972	0.990	0.981
S	HSS	7.69	28.13	6474.01	0.970	0.972	0.990	0.986
S	HSP	3.08	43.42	1415.14	0.971	0.972	0.990	0.990
S	SSS	0.07	10.88	1.48	0.971	0.974	0.989	0.983
S	SSP	0.76	39.50	235.35	0.971	0.973	0.989	0.991
S	SPP	0.02	1.74	133.67	0.972	0.973	0.993	0.990
S	PPP	0.06	0.54	51.58	0.972	0.980	0.993	0.989

[1] σ^2_H: Observation variance, herd effect, σ^2_S: Observation variance, section effect, σ^2_H: Observation variance, pen effect
[2] δ_1: LG, δ_2: H1, δ_3: H2, δ_4: H3

variance tradeoff describes how a model either adjusts too well to the training data, hereby reducing the bias (MSE) and increasing the variance, or decreases the variance by a reduced sensitivity to the learning data, which results in a higher bias (Torgo, 2017; Witten and Frank, 2005). Since the estimation of variance components in the presented DLM were aiming for the lowest MSE, and the variables in the model are highly correlated, it is very likely, that the model is overfitting the learning data.

Although overfitting in general is sought avoided and does not add to an increased performance (Hawkins, 2004), it does not necessarily impede the predictive performance either (Elith et al., 2008; Lieberman and Morris, 2014). Overfitting is a well known challenge when handling *correlated* variables or variables analyzed using *regression* techniques (Hawkins, 2004). But Elith et al. (2008) found that the predictive performance of an overfitted model was unaffected when using *Boosted Regression Trees* and evaluating using cross-validation. Likewise Lieberman and Morris (2014) concluded that multi-collinearity in cross-validated models, was irrelevant if prediction performance was the goal of the model.

6.5.2 *Predictive performance*

In Tables 6.5 and 6.6 the MSE of both learning data and test data from each herd is presented. The predictive performance of a model should not be evaluated on learning data (Elith et al., 2008), but we still choose to present the models' MSE on learning data in order to show how the MSE is lower on learning data due to possible overfitting, as described above.

When comparing the MSE of the learning data to the MSE of the test data for Herd A, it can be seen, that the MSE of the test data is 3.5 times the MSE of the learning data on average (see

Table 6.5: MSE for learning data and for test data for seven model structures for Herd A. Notations: LG = Linear Growth model, H1 = Cyclic model of length 24, H2 = Cyclic model of length 12, H3 = Cyclic model of length 8. H = Herd level, S = Section level, P = Pen level

Model Structure		MSE	
LG	H1 H2 H3	Learning data [1]	Test data [2]
S	HHH	4.003	15.687
S	HSS	3.923	14.535
S	HSP	3.921	14.612
S	SSS	3.954	13.850
S	SSP	3.850	13.976
S	SPP	3.836	13.946
S	PPP	3.800	13.924

[1] 9540 hours (68 %)
[2] 4441 hours (32%)

Table 6.5), whereas the MSE on learning data is only 1.3 times higher than for test data for Herd B on average (see Table 6.6). A larger difference between MSE on learning data and test data could be expected for Herd B because of the very large observational variances and low discount factors previously discussed. However, the ratio between the two MSE values is not relevant when choosing the model version with the better predictive performance. In the following comparison of model versions, we will only refer to the MSE of the test data.

6.5.3 *Model versions - Herd A*

When comparing the MSE of the different model versions from Herd A (see Table 6.5), it shows that all model versions which include cyclic sub models defined at herd level have the highest MSE. This indicates that differences between pens in the herd are too large to be described by the same cyclic sub model.

This is well illustrated in Figure 6.8, which shows the drinking pattern of a week in four pens in four sections in Herd A. The diurnal drinking pattern in pen 1.6 is disturbed, which results in the water consumption peaking every eight hour instead of once per 24 hours as in the undisturbed diurnal patterns of pens 2.5 and 3.5. This abnormal pattern in pen 1.6, in combination with a longer sensor outage in pen 5.7, decreases any correlation between sections, and situations like this are likely to cause the herd level model versions to under perform.

The planned periods of missing data during cleaning periods between batches are likely to affect the model versions with parameters defined at herd level as well. Our initial assumption was that a distinct diurnal pattern would characterise pigs' drinking pattern throughout the growth period. However, the pigs in pen 1.6 were inserted 10 weeks before the depicted week, and showed no such diurnal drinking pattern at this time.

Table 6.6: MSE (mean squared error) for learning data and for test data for seven model structures for Herd B. Notations: LG = Linear Growth model, H1 = Cyclic model of length 24, H2 = Cyclic model of length 12, H3 = Cyclic model of length 8. H = Herd level, S = Section level, P = Pen level

Model Structure		MSE	
LG	H1 H2 H3	Learning data [1]	Test data [2]
S	HHH	1.350	1.750
S	HSS	1.327	1.727
S	HSP	1.325	1.712
S	SSS	1.235	1.621
S	SSP	1.234	1.559
S	SPP	1.229	1.556
S	PPP	1.194	1.466

[1] 14657 hours (83 %)
[2] 3025 hours (17%)

Such a disturbed diurnal pattern indicates that some pigs have to drink during the night time in order to get their need for water satisfied. Drinking activity in finisher pigs during the night was also found by Andersen et al. (2016), and installing an extra drinking nipple might be necessary to restore the diurnal pattern, and supply sufficient amounts of water to the pigs. This is especially profound for finisher pigs that are fed liquid feed, since the restrictively feeding from 60 kg reduces the amount of water assigned through the feed and increases the demand for water from the drinking nipples.

The results in Table 6.5 also show that the model version with all four sub models defined at section level outperforms any model version including cyclic sub models defined at pen level. This supports the initial hypothesis of a correlation between pens and sections in a herd, which can be modeled in a spatial model.

6.5.4 *Model versions - Herd B*

In Table 6.6 we see that the best performing model version is the one with all cyclic sub models defined at pen level. This result indicates that any correlation between pens and sections in a herd may be described solely through the correlation structure of the system variance-covariance matrix, $\mathbf{W}_t$, as described in Section 6.3.3 and through the observation variance-covariance matrix described in Section 6.3.3.

In general, model versions which included cyclic sub models at section level, performed better than any model version including herd level sub models. The herd level versions may fail in Herd B for the same reasons as in Herd A, although the cleaning periods between batches were shorter in Herd B.

The the poorer performance of models with higher degrees of spatial correlation is surprising considering the very disciplined sectionized management on the research farm and the uniformity

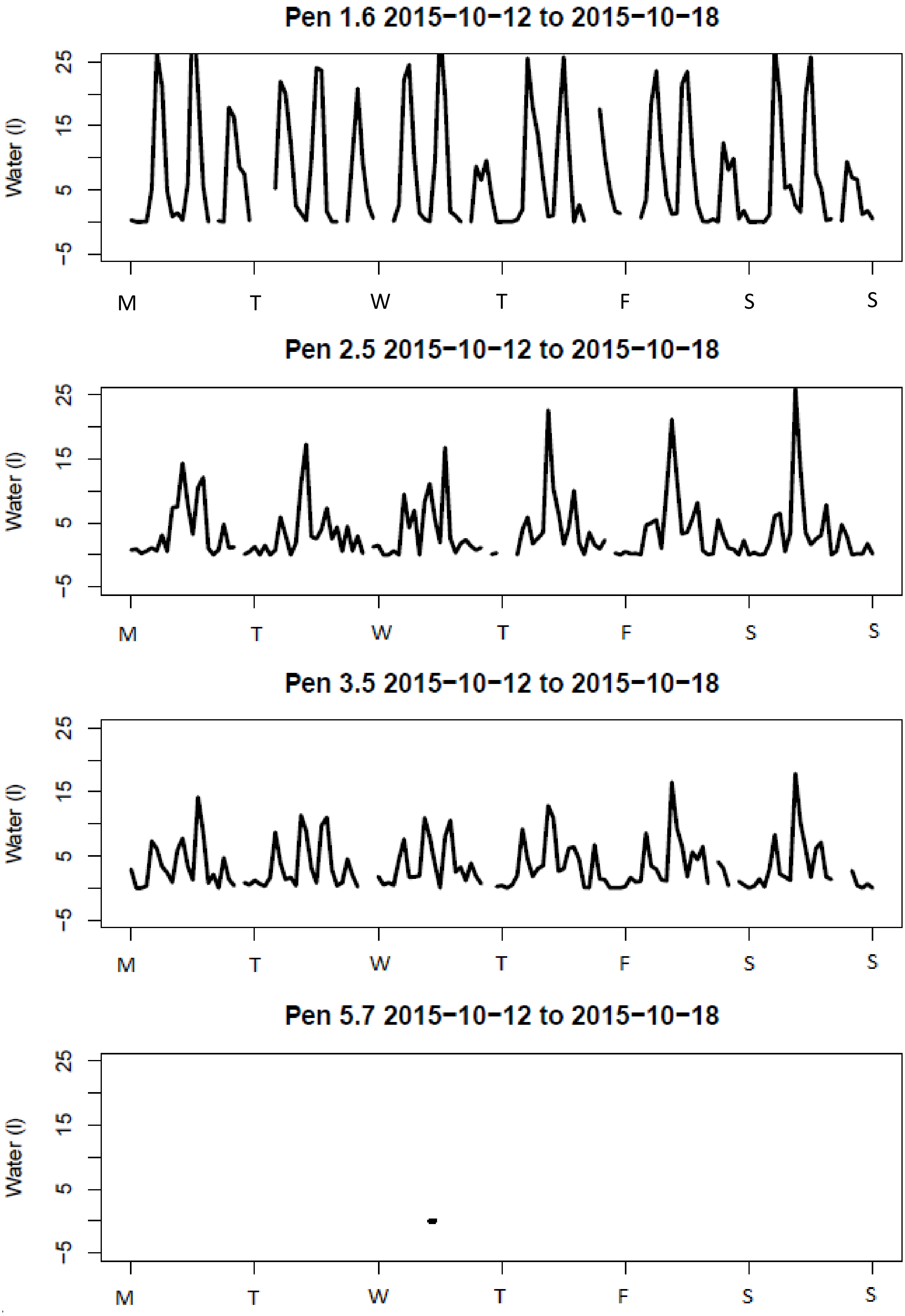


Figure 6.8: The water consumption from four pens in four different sections of Herd A in the same week. The pigs in each of the four pens are of different ages with the oldest in pen 1.6. The diurnal pattern is disturbed in pen 1.6, but intact in pens 2.5 and 3.5. In pen 5.7 there is no data due to sensor outage.

of the pigs both within sections and within the herd relative to a finisher herd. The main reason for the result may be the overfitting of the model as previously discussed. Since overfitting models tend to model the random noise of a system, as described by Witten and Frank (2005), they are more flexible and quickly adjust to smaller changes (Torgo, 2017). This is in compliance with the results of the estimation of variance components. An adjustment to smaller changes can cause the model to emphasize the importance of random noise within the single pen, hence failing to recognize any correlation between pens.

6.5.5 *Estimation procedure*

Since the end goal of the project is to use the developed DLMs for early warning about undesired events it could be argued that the discount factors and observation variances should have been estimated on a learning data set where such events did not occur at all. Such an approach was, for instance, used by Jensen et al. (2017) and the obvious advantage would be that if parameters are estimated under “normal” conditions, deviations from the normal pattern in case of undesired events are more easily detected.

In this study, the learning data set was simply the first batches of the study period and the test data set was the last two batches. The reason for this choice was that we present a framework for simultaneous monitoring of the entire herd. Thus, it is not possible to find batches without any undesired events in any pen of the herd as it was for Jensen et al. (ibid.) who basically modeled a single pen. Thus, the advantage of fitting the model to “normal” conditions is lost which, potentially, will make it more difficult to detect deviating data patterns caused by undesired events.

An indirect estimation technique through discount factors was used for assessing the system variance-covariance matrices of the models. Several initial attempts were done to estimate all variance components more directly by the EM algorithm (see for instance Dethlefsen, 2001) as previously done in multivariate DLMs by Bono et al. (2012) and Jensen et al. (2017) but the iterative algorithm failed to converge and some of the variances drifted out of scope over the iterations. Therefore, the indirect approach with discount factors was used instead.

The full univariate model (i.e. for one sensor) as presented in Section 6.3.3 is directly inspired by the work of Madsen et al. (2005) who concluded that a super positioned model consisting of a linear growth component and three harmonics described the diurnal pattern well. It is interesting that Madsen (2001, Chapter 8) reported that the EM algorithm also failed for the univariate model despite several attempts. Thus, it seems to be a pattern that the EM algorithm is not well suited for estimation in models with harmonics based on Fourier form representation of seasonality. The observation is supported by unpublished work by the authors in relation to similar models with diurnal patterns.

Madsen (ibid.) mentioned that the system variance-covariance matrix must be expected to change over time as the pigs grow. Therefore, it was argued that an “online” estimation technique based on discount factors should actually be preferred to estimation by the EM algorithm. It is not clear whether the system variance-covariance actually do change over time, but if Madsen (ibid.) is right, the discount factor approach will also be a good choice for the present study.

It was also argued that the direct link to the variability of data as expressed by Eqs. (22) and (23) makes the approach less herd specific, because no new estimation is needed for each herd (as long as the estimated discount factors are valid across herds). The results from this study (cf. Tables 6.3 and 6.4) do not confirm that discount factors are identical for different herds (at least not for a multivariate spatial model) so it is expected that an estimation step is needed for each herd before the system is ready for use.

When looking at the observation variances of Tables 6.3 and 6.4 it is quite obvious that the values are *not* estimates for the true observation variances. They are far too big and they should just be seen as the values optimizing the fit in combination with the resulting discount factors. The very big values seen for many of the models have as a consequence that the models become less adaptive which may actually be an advantage given that the ultimate use of the models is to produce early warnings of deviating patterns caused by undesired events.

A similar behavior could have been achieved with discount factors closer to 1, but the best fit was apparently achieved by very high "observational" variance and smaller discount factors. It is, however a question whether a generalized (to the multivariate case) version of the Kalman filter with unknown observational variance should have been developed for this study. Madsen et al. (2005) used that approach for the univariate case so a suggestion for future research is to see whether the method can be extended to the multivariate case.

Even though the observational error structure described in Eqs. (18) and (19) intuitively seems natural, it can be discussed whether it actually over-parameterizes the observation errors. Particularly, it is a question whether the herd level variance (σ_H^2) should have been taken out. For Herd A, this term only contributed with up to 5% of the total observation variance and for Herd B it never even reached 1% of the variance (percentages calculated from Tables 6.3 and 6.4. Thus, in future studies, it is recommended not to include the herd level variance.

6.6 CONCLUSION

We can conclude, that it is possible to develop a spatial DLM for the modeling of drinking patterns across a herd of growing pigs. In Herd A, the model version expressing the strongest correlation in the drinking patterns between pens within a section (the SSS model) obtains the highest fit. Model versions which include parameters at pen level (SSP, SPP, and PPP) fit almost as well, whereas model versions with parameters expressed at herd level (HHH, HSS, and HSP) fit the worse. Based on model fit, correlation between pens do occur in Herd A, but primarily between pens within the same section. In Herd B, a distinctive inverse relation between model fit and degree of correlation in the drinking patterns are found. This results in the model version with the least correlation between the pens (the PPP model) obtaining the highest fit, and the model version with the highest degree of correlation (the HHH model) obtaining the poorest fit. Thus, based on model fit, little or no correlation between pens occur in Herd B. For both herds, overfitting of test data may influence the results.

6.7 PERSPECTIVES

The overall motivation for the development of the presented model is to investigate spatial modeling of water consumption as a strategy for a future detection system in commercial pig production. However, the ability of the model to identify unwanted events in the herd must be evaluated before taking any further steps. Such an evaluation of the detection performance will be conducted in a following paper. Thus, the forecast errors generated by the seven model versions will be monitored in a control chart, and the ability of the detection system to predict outbreaks of either diarrhea or fouling is described in (Dominiak et al., 2017).

6.8 ACKNOWLEDGEMENTS

This research was funded by the Danish Council for Strategic Research (The PigIT project, Grant number 11-116191). We further wish to thank the anonymous farmers, the Danish Pig Research Centre and the technical staff (particularly Mads Ravn Jensen) at Aarhus University for installing and supervising the sensors and for taking care of the daily observations in the herd.

CHAPTER 6 REFERENCES

Andersen, H. M. L., E. Jørgensen, and L. J. Pedersen (2016). „Using Evolutionary Operation technique to evaluate different management initiatives at herd level.“ In: *Livestock Science* 187, pp. 109–113.

Anonymous (2000). *Instruction Leaflet Flow sensors - 15 mm Dia. pipe V8189*. URL: http://docs-asia.electrocomponents.com/webdocs/001b/0900766b8001bb47.pdf.

Berckmans, D. (2014). „Precision livestock farming technologies for welfare management in intensive livestock systems. (Special Issue: Animal welfare: focusing on the future.)“ In: *Revue Scientifique et Technique - Office International des Epizooties* 33.1, pp. 189–196.

Bono, C., C. Cornou, and A. R. Kristensen (2012). „Dynamic production monitoring in pig herds I: modeling and monitoring litter size at herd and sow level.“ In: *Livestock Science* 149.3, pp. 289–300.

Bressers, H. P. M., J. H. A. Te Brake, and J. P. T. M. Noordhuizen (1995). „Automated oestrus detection in group-housed sows by recording visits to the boar“. In: *Livestock Production Science* 41.2, pp. 183–191.

Cameron, R. D. A. (2000). „A review of the industialisation of pig production worlwide with particular reference to the asian region“. In: *Food and Agriculture Organization of the United Nations*.

URL: {http://www.fao.org/ag/againfo/resources/en/publications/agapubs/awi_concept_pig_product.pdf}.

Cavero, D. et al. (2006). „Mastitis detection in dairy cows by application of fuzzy logic.“ In: *Livestock Science* 105.1/3, pp. 207–213.

Cavero, D. et al. (2007). „Analysing serial data for mastitis detection by means of local regression.“ In: *Livestock Science* 110.1/2, pp. 101–110.

Cornou, C. and S. Lundbye-Christensen (2008). „Classifying sows' activity types from acceleration patterns: an application of the Multi-Process Kalman Filter.“ In: *Applied Animal Behaviour Science* 111.3/4, pp. 262–273.

Danish Agriculture and Food Council (2010). *Danish Pig Producers and Animal Welfare*. Tech. rep. Danish Agriculture and Food Council.

Dethlefsen, C (2001). „Space Time Problems and Applications“. PhD thesis. Aalborg University.

Dominiak, K. N. and A. R. Kristensen (2017). „Prioritizing alarms from sensor-based detection models in livestock production - a review on model performance and alarm reducing methods.“ In: *Computers and Electronics in Agriculture* 133, pp. 46–67.

Dominiak, K. N. et al. (2017). „Reducing alarms and prioritising interventions in pig production by simultaneous monitoring of water consumption in multiple pens“. In: *8th European Conference on Precision Livestock Farming*.

Elith, J., J. R. Leathwick, and T. Hastie (2008). „A working guide to boosted regression trees“. In: *Journal of Animal Ecology* 77.4, pp. 802–813. DOI: 10.1111/j.1365-2656.2008.01390.x.

Freson, L. et al. (1998). „Validation of an infra-red sensor for oestrus detection of individually housed sows.“ In: *Computers and Electronics in Agriculture* 20.1, pp. 21–29.

Garcia, E. et al. (2014). „Lameness detection challenges in automated milking systems addressed with partial least squares discriminant analysis.“ In: *Journal of dairy science* 97.12, pp. 7476–7486.

Hawkins, Douglas M. (2004). „The Problem of Overfitting“. In: *Journal of chemical information and computer sciences* 44.1. doi: 10.1021/ci0342472 pmid: pp. 1–12.

Hertem, T. van et al. (2013). „Lameness detection based on multivariate continuous sensing of milk yield, rumination, and neck activity.“ In: *Journal of dairy science* 96.7, pp. 4286–4298.

Hertem, T. van et al. (2014). „Automatic lameness detection based on consecutive 3D-video recordings.“ In: *Biosystems Engineering* 119, pp. 108–116.

Hogeveen, H. et al. (2010). „Sensors and clinical mastitis - the quest for the perfect alert.“ In: *Sensors* 10.9, pp. 7991–8009.

Huybrechts, T. et al. (2014). „Early warnings from automatic milk yield monitoring with online synergistic control“. In: *Journal of dairy science* 97.6, pp. 3371–3381.

Jensen, D. B., N. Toft, and A. R. Kristensen (2017). „A multivariate dynamic linear model for early warnings of diarrhea and pen fouling in slaughter pigs.“ In: *Computers and Electronics in Agriculture* 135, pp. 51–62.

Kamphuis, C. et al. (2010). „Decision-tree induction to detect clinical mastitis with automatic milking.“ In: *Computers and Electronics in Agriculture* 70.1, pp. 60–68.

Kamphuis, C., J. K. Burke, and J. G. Jago (2013). „Cows becoming clinically lame differ in changes in behaviour and physiology compared to cows that do not become clinically lame.“ In: *Proceedings of the New Zealand Society of Animal Production* 73, pp. 5–10.

Kashiha, M. et al. (2013). „The automatic monitoring of pigs water use by cameras.“ In: *Computers and Electronics in Agriculture* 90, pp. 164–169.

Kramer, E. et al. (2009). „Mastitis and lameness detection in dairy cows by application of fuzzy logic.“ In: *Livestock Science* 125.1, pp. 92–96.

Kristensen, A. R., E. Jørgensen, and N. Toft (2010). *Herd Management Science. I. Basic concepts*. Copenhagen: Academic books. ISBN: 9788763461191.

Krogsdahl, J. (2014a). *Description of Grønhøj Farm.* PigIT Report 6. Department of Large Animal Sciences, University of Copenhagen.: The PigIT project.

Krogsdahl, J. (2014b). *Description of Kappel Farm*. PigIT Report 5. Department of Large Animal Sciences, University of Copenhagen.: The PigIT project.

Lieberman, M. and J. D. Morris (2014). „The Precise Effect of Multicollinearity on Classification Prediction“. In: *Multiple Linear Regression Viewpoints* 40.1, pp. 5–10.

Maatje, K., R. M. de Mol, and W. Rossing (1997). „Cow status monitoring (health and oestrus) using detection sensors“. In: *Computers and Electronics in Agriculture* 16.3, pp. 245–254.

Madsen, T. N. (2001). „Tools for monitoring growing pigs“. PhD thesis. The Royal Veterinary and Agricultural University.

Madsen, T. N., S. Andersen, and A. R. Kristensen (2005). „Modelling the drinking patterns of young pigs using a state space model.“ In: *Computers and Electronics in Agriculture* 48.1, pp. 39–62.

Maertens, W. et al. (2011). „Development of a real time cow gait tracking and analysing tool to assess lameness using a pressure sensitive walkway: the GAITWISE system.“ In: *Biosystems Engineering* 110.1, pp. 29–39.

Mol, R. M. de and W. Ouweltjes (2001). „Detection model for mastitis in cows milked in an automatic milking system.“ In: *Preventive veterinary medicine* 49.1/2, pp. 71–82.

Mol, R. M. de et al. (1997). „Results of a multivariate approach to automated oestrus and mastitis detection“. In: *Livestock Production Science* 48.3, pp. 219–227.

Mol, R. M. de et al. (1999). „Description of a detection model for oestrus and diseases in dairy cattle based on time series analysis combined with a Kalman filter.“ In: *Computers and Electronics in Agriculture* 22.2/3, pp. 171–185.

Ostersen, T., C. Cornou, and A. R. Kristensen (2010). „Detecting oestrus by monitoring sows' visits to a boar“. In: *Computers and Electronics in Agriculture* 74.1, pp. 51–58.

Pastell, M. E. and M. Kujala (2007). „A Probabilistic Neural Network Model for Lameness Detection“. In: *Journal of dairy science* 90.5, pp. 2283–2292.

Pedersen, Ken S. (2012). „Smaagrisediarre“. In: *Dansk Veterinaer Tidsskrift* 6, pp. 18–22.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `http://www.R-project.org/`.

Rasmussen, M. D. (2002). „Defining acceptable milk quality at time of milking“. In:

Sorensen, C. G. et al. (2010). „Conceptual model of a future farm management information system.“ In: *Computers and Electronics in Agriculture* 72.1, pp. 37–47.

Torgo, Luís (2017). *Data Mining with R Learning with Case Studies*. Chapman & Hall.

Viazzi, S. et al. (2013). „Analysis of individual classification of lameness using automatic measurement of back posture in dairy cattle“. In: *Journal of dairy science* 96.1, pp. 257–266.

Vils, E. (2012). *Vaadfoderkurver*. Tech. rep. SEGES.

Vils, Else et al. (2013). *Manal om vaadfodermanagement*. Tech. rep. SEGES, Danish Pig Research Centre.

West, Mike and Jeff Harrison (1999). *Bayesian Forecasting and Dynamic Models*. 2nd. Springer.

Witten, I.H. and E. Frank (2005). *Data mining: practical machine learning tools and techniques*. 2nd edition. Morgan Kaufmann, Elsevier.

7

PAPER III

Submitted for publishing in *Computers and Electronics in Agriculture*: Dominiak et al. (2017b)

SPATIAL MODELING OF PIGS' DRINKING PATTERNS AS AN ALARM REDUCING METHOD
II. APPLICATION OF A MULTIVARIATE DYNAMIC LINEAR MODEL

K. N. Dominiak, J. Hindsborg, L.J. Pedersen, and A. R. Kristensen

Abstract: The objectives of this paper are to evaluate the detection performance of a previously developed multivariate spatial dynamic linear model (DLM), and to discuss potential post processing strategies for reducing alarms. Performance evaluation is conducted by applying a *standardized tabular two-sided Cusum* to the forecast errors generated by the spatial model. For two herd, the forecast errors are generated at pen level, section level, and herd level. Seven model versions express different degrees of correlations in the drinking patterns between pens and sections in a herd, and the performances of the three spatial levels are evaluated for each of the model versions. The alarms generated by the Cusum are categorized as true positive (TP), false positive (FP), true negative (TN), or false negative (FN) based on time windows of three different lengths. In total, 126 combinations of herds, spatial levels, model versions, and time windows are evaluated, and the performance of each combination is reported as the *area under the ROC curve* (AUC). The highest performances were obtained when detecting events at herd level (AUC = 0.98 (weaners) and 0.94 (finishers)) using the longest time window and expressing the highest degree of correlation. However, the settings most suitable for implementation in commercial herds, were obtained when detecting events at section level (AUC = 0.86 (weaners) and 0.87 (finishers)) using the medium-length time window and highest degree of correlation. The combination of a spatial DLM and a two-sided Cusum has high potential for prioritizing high-risk alarms as well as for merging alarms from multiple pens within the same section into a reduced number of alarms communicated to the caretaker. Thus, the spatial detection system described in this paper and in a previous paper constitute a new and promising approach to sensor based monitoring tools in livestock production.

7.1 INTRODUCTION

For more than 20 years the development of sensor-based detection models within the field of livestock science has been subject to an increasing scientific focus. However, a general challenge for detection models is that they generate too many false alarms (Dominiak and Kristensen, 2017; Hogeveen et al., 2010). False alarms reduce the reliability of a detection model as a decision-support tool, and represent a major reason for models being unsuited for implementation in modern livestock production herds (Hogeveen et al., 2010; Mein and Rasmussen, 2008).

Livestock science detection models are generally designed to detect individual animals having specific conditions, like oestrus or clinical mastitis (CM), which are relatively rare when compared to animals not having the condition (Huybrechts et al., 2014; Kamphuis et al., 2010; Ostersen et al., 2010). Correct identification of rare conditions increases the demands for a high specificity if too many false alarms should be avoided (Dominiak and Kristensen, 2017; Hogeveen et al., 2010; Rasmussen, 2002).

The amount of false positive (FP) alarms can be reduced through improved performance of the detection models or through a post processed prioritization, or sorting, of the generated alarms into true or false. Sufficient performance is the first of three success criteria for implementation of CM detection models in commercial farms, described by Hogeveen et al. (2010). The three criteria are **a**) a sensitivity (Se) of minimum 80% and a specificity (Sp) of minimum 99%, **b**) a detection window of maximum 48 hours, and **c**) the circumstances of the study must be as similar to practical farm circumstances as possible. Following these criteria, no sensor-based detection models in livestock production (1995-2015) are suitable for implementation (Dominiak and Kristensen, 2017). Although the criteria may not fully apply to other conditions than CM, they can still serve as a guideline for evaluating a model's ability to be implemented in commercial herds.

In scientific literature, only three alarm prioritizing methods are described (ibid.), of which Naïve Bayesian Network (NBN) seems the most suitable for postprocessing output from different types of modeling methods. An alternative approach to detecting conditions at individual animal level, is a spatial approach, which aims to identify a specific area of the production unit, which needs managerial attention. Danish pig producing units for growing pigs (weaners 7-30 kg and finishers 30-110 kg) are very well suited for spatial modeling. In order to maintain a high level of bio-security, these units are run with a clear spatial separation between pigs of different age groups (Danish Agriculture and Food Council, 2010).

Such a construction of the herd enables a spatial modeling of a production site as one whole production unit (the herd) consisting of a number of identical subunits (sections) each consisting of a number of identical sub-subunits (pens). A warning system based on a spatial modeling of the production unit, can communicate area-specific warnings, which will provide the manager with valuable information when choosing an intervention to prevent or reduce the consequences of the condition.

In this paper we will present the second, and final, part of the full description of a spatial approach to the challenge of reducing alarms. The description is initiated in Dominiak et al. (2017a) where the development of a multivariate dynamic linear model (DLM) for modeling water consumption of growing pigs in Denmark is described. The structure of the DLM will be described in short terms in Section 7.2. The objectives of this paper are to evaluate model performances on test data from two different herds, using Area Under Curve (AUC), and to discuss post processing strategies for reducing the number of alarms that are generated.

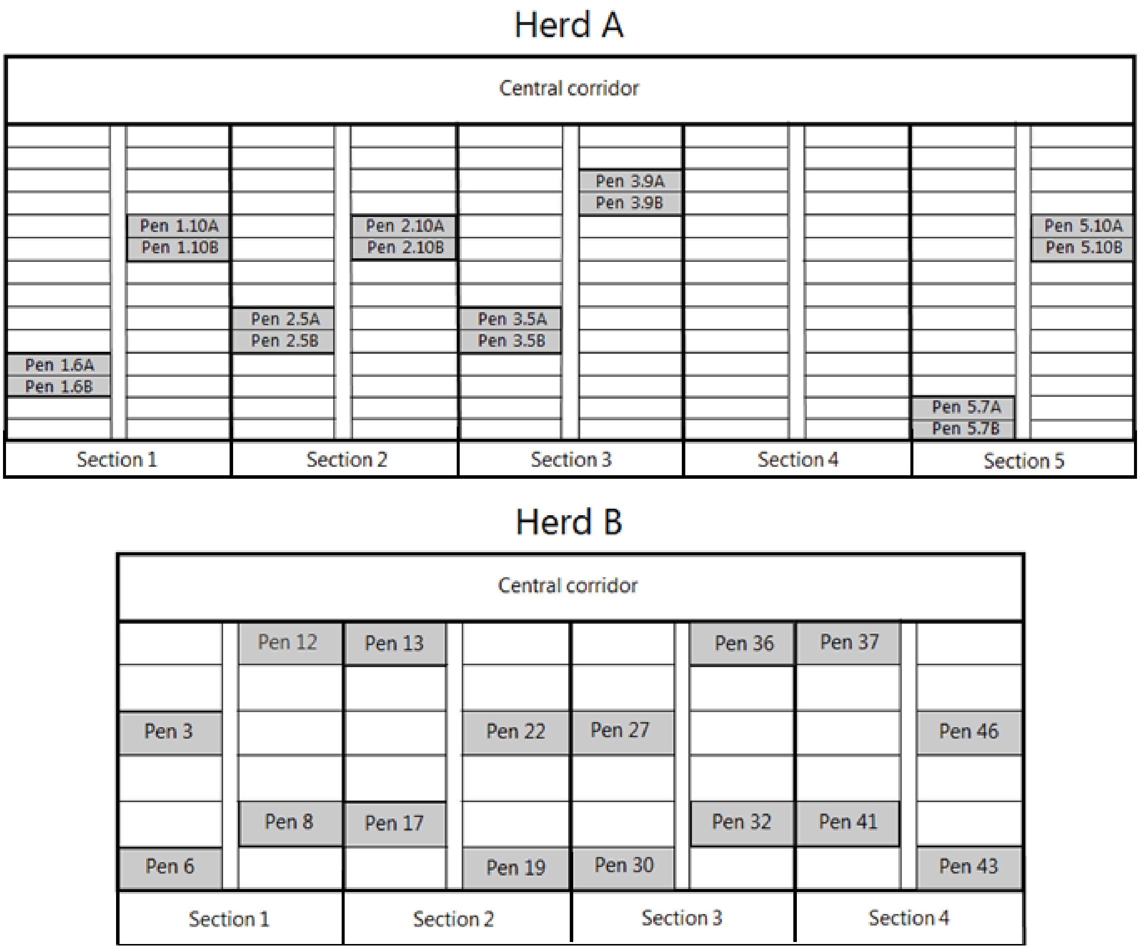


Figure 7.1: Structure of Herd A and Herd B. Grey double pens in Herd A and grey pens in Herd B were equipped with sensors and included in the study. From Dominiak et al. (2017a).

7.2 HERDS, DATA AND MODELS

In this section a short background is given to the spatial DLM whose predictive accuracy is evaluated in the present paper. It is based on Dominiak et al. (ibid.), where a detailed description of the model development can be found.

7.2.1 *Herd description*

Water consumption data for this study was obtained from two different herds. Herd A is a Danish commercial finisher herd, and Herd B is an experimental weaner herd, “Grønhøj”, which is owned by the Danish Pig Research Centre.

Herd A produces 10.000 cross-bred finisher pigs per year, and the herd has five identical sections, of which four are included in this study (Figure 7.1 (A)). Each section consists of 28 pens, where 18 pigs are inserted per pen. Two neighbouring pens share the same water pipe, which supplies one drinking nipple in each of the two pens (36 pigs).

Table 7.1: Characteristics for the two herds in the study (14 for section K11). From Dominiak et al. (2017a).

Characteristic	Herd A	Herd B
Production type	Commercial	Research Farm
Animal group	Finishers (30-110 kg)	Weaners (7-30 kg)
Sections	4	4
Sensors total/ per section	8/2	16/4
Pigs per pen/ per sensor	18/36	15/15
Growth period (batch)	14 weeks	8 weeks
Batches per sensor	7	13 [1]
Learning data (hours)	9540	14657
Test data (hours)	4441	3025

[1] 14 for section 4.

Herd B consists of four sections, each with 12 pens for weaner pigs, where 15 pigs are inserted per pen (Figure 7.1 (B)). Each pen is supplied by individual water pipes, hence one water pipe supplies one drinking bowl per pen (15 pigs).

7.2.2 *Sensor data*

Water consumption data was obtained by a flow meter (RS V8189 15mm Diameter Pipe) (Anonymous, 2000), which was placed on the water pipe supplying either drinking nipples (Herd A) or bowls (Herd B) in the pens. The data was converted to litres before it was aggregated per hour, yielding water use in litres per hour as input from each sensor to the DLM.

A total of eight sensors were installed in Herd A with two sensors in each of four sections, and each sensor monitoring the water consumption of two neighbouring pens (36 pigs). Data from one sensor creates an individual time series, hence the full data set from Herd A consists of eight time series, or variables, which are monitored simultaneously.

Sixteen sensors in total were installed in Herd B with four sensors in each of four sections. Each sensor monitored the water consumption of one single pen (15 pigs) in individual time series. The full data set from Herd B therefore consists of sixteen time series, or variables, which are monitored simultaneously.

The main characteristics of the two herds are summarized in Table 7.1.

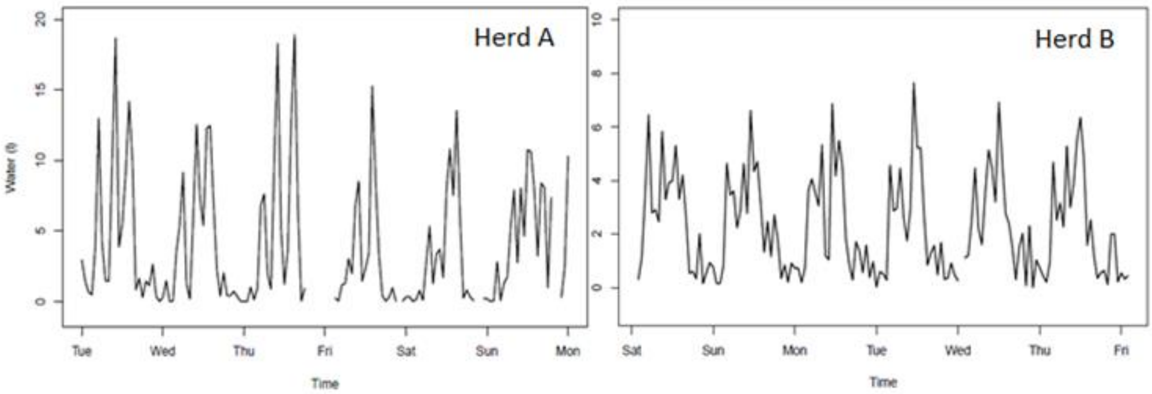


Figure 7.2: Diurnal drinking pattern of finishers (Herd A) and weaners (Herd B). From Dominiak et al. (2017a).

7.2.3 *Modeling drinking patterns*

The drinking patterns of both weaners and finishers follow a clear diurnal pattern (Figure 7.2), and the underlying level of water consumed increases over time, indicating that pigs drink more as they grow (Madsen et al., 2005).

The diurnal drinking patterns can be described by the sum of four dynamic linear models; three for harmonic waves (H1, H2, H3) and one for the underlying linear growth (LG), which are superpositioned into the final full DLM (ibid.).

The amount of water consumed within the last hour at time t for each of the n sensors is expressed in the observation vector $Y_t = (Y_{1t}, \ldots, Y_{nt})'$. The aim of the DLM is to predict the next observation by estimating the parameter vectors $\theta_1, \ldots, \theta_t$ from the observations.

The accuracy of the predictions is expressed through forecast errors e_t, which contain any difference between the predicted observation and the actual observation. As long as the drinking pattern reflects a normal situation and evolves as expected, the prediction of the next observation is close to perfect, and any forecast error will be small. Should the pigs, for some reason, drink more or less than expected, the predictions and the observations will diverge, and the errors will be larger. A systematic change in the normal drinking pattern will generate a sequence of forecast errors, which will lead to an alarm when plotted in a control chart, as described in Section 7.3.

7.2.4 *Model versions*

Each of the four sub-models can be defined at herd, section or pen level to allow for the diurnal pattern to evolve differently between pens or sections in the herd. Seven different model versions were defined (Table 7.2) with the LG model defined at section level in all versions due to the equality in size and age of pigs within a section.

Table 7.2: Model versions applied to data sets from Herd A and Herd B. The Linear Growth sub model is defined at Section level in all models, whereas different combinations of level definitions are made for the cyclic sub models. Notations: LG = Linear Growth model, H1 = Cyclic model of length 24, H2 = Cyclic model of length 12, H3 = Cyclic model of length 8. H = Herd level, S = Section level, P = Pen level. From Dominiak et al. (2017a).

LG	H1 H2 H3	Interpretation
S	HHH	The full harmonic pattern evolves identically for all pens in the herd
S	HSP	H1 evolves identically for all pens, H2 evolves identically within sections but differently between sections, H3 evolves differently in each pen
S	HSS	H1 evolves identically for all pens, H2 and H3 evolve identically within each section but differently between sections
S	SSS	The full harmonic pattern evolves identically within each section but differently between sections
S	SSP	H1 and H2 evolve identically within sections but differently between sections, H3 evolves differently in each pen
S	SPP	H1 evolves identically within sections but differently between sections, H2 and H3 evolve differently in each pen
S	PPP	The full harmonic pattern evolves differently in each pen

7.2.5 *Model output*

From each of the seven model versions a series of forecast error vectors (e_t) and a series of forecast variance-covariance matrices (Q_t) are generated. The forecast errors and variances are entered as input variables to a *standardized two-sided Cusum control chart*, as described by Montgomery (2013). Systematic changes in the water consumption will then generate alarms that are evaluated as an expression of the predictive performance of the model versions.

7.3 EVALUATING MODEL PERFORMANCE

7.3.1 *Events of interest*

The events of interest in this study are diarrhea and fouling amongst growing pigs. Both diarrhea and fouling, which is a change in behaviour where the pigs start to lie on the slatted area of the pen and excrete in the lying area, reduce animal welfare (Aarnink et al., 2006; Pedersen, 2012). Every morning, the caretakers at each farm register if either of the two events occur in a pen or not. The routines for assessment of either event is described in a project protocol (Lyderik et al., 2016), and are calibrated by an experienced technician regularly throughout the study period.

If considered necessary by the caretaker, pigs with diarrhea are treated with antibiotics, and pens where fouling has occurred are cleaned. Event registration and treatments are conducted once a day, but because the actual outbreak of the event can happen at any hour between two registrations, an event is defined to last 24 hours from midnight to midnight in the present study.

The objective of this paper is, however, to evaluate model performances at different spatial levels rather than the ability to distinguish between specific conditions. Therefore registrations of both diarrhea and fouling are joined under the common term “event”. This more general definition of events is supported by Madsen et al. (2005) and Andersen et al. (2016), stating that changes in drinking patterns can reflect changes to the general wellbeing of pigs. Changes may therefore not be uniquely related to a specific type of event.

Despite regular calibration of registration routines, significant herd-specific differences in the frequency of event registrations occur, and two different event definitions are used: In Herd A the daily caretaker was replaced with unexperienced personnel a number of times during the period of data collection. As a consequence of that, the commitment to register daily events was inconsistent and resulted in periods with no registrations. For performance evaluation on Herd A data, all event registrations available for the herd constitute the gold standard.

In Herd B the threshold for identification of diarrhea was low. This lead to multiple periods with registrations of diarrhea every day for 14-21 days, although only few or no interventions were made during those periods. For performance evaluation on Herd B data, the initiation of an intervention (medical treatment of diarrhea or cleaning of pens with fouling), rather than daily event registrations, constitute the gold standard.

7.3.2 *Time Window*

By comparing alarms and events occurring at the same moment, the alarms can be categorized true or false, and the performance of the model can be calculated (Hogeveen et al., 2010). But alarms seldom occur at the exact same moment as the events, and if they did, they were of little predictive value. Therefore *time windows* are often used (Hogeveen et al., 2010; Jensen et al., 2017; Ostersen et al., 2010). A time window is a defined period of time associated with a registered event, and any number of alarms occurring within that window is treated as one single alarm, and categorized as detecting the event correctly (Figure 7.3).

Time windows can be of varying lengths, and can extend from before an event to after an event (Jensen et al., 2017; Mol et al., 1997). As Figure 7.3 illustrates, the length of a time window has great influence on the categorization of true or false alarms, and therefore on the performance of a model (Hogeveen et al., 2010). Longer windows improve model performance, whereas windows extending beyond an event can result in alarms being communicated after the event has occurred. The categorization of alarms as true or false are counted as follows:

- Alarms within a time window are counted as one *true positive* (TP)
- Alarms occurring outside of a time window are counted as *false positive* (FP)
- If no alarms occur within a time window, it is counted as *false negative* (FN)
- Days without alarms and with no time window are counted as *true negative* (TN)

The detection accuracy can then be expressed by *sensitivity* (Se) and *specificity* (Sp), which are calculated as:

$$Se = \frac{TP}{(TP + FN)} \tag{24}$$

and

$$Sp = \frac{TN}{(TN + FP)} \tag{25}$$

where TP denotes the total number of TP cases and accordingly for the other variables.

Three lengths of time windows are applied for the performance evaluation in this paper (see Figure 7.3). The longest window includes three days before an event plus the day of the event, but zero days after. The two other window lengths includes two days and one day before an event respectively plus the day of the event, but none after. The three windows are denoted (3/0), (2/0), and (1/0) respectively, following the terminology of Jensen et al. (2017).

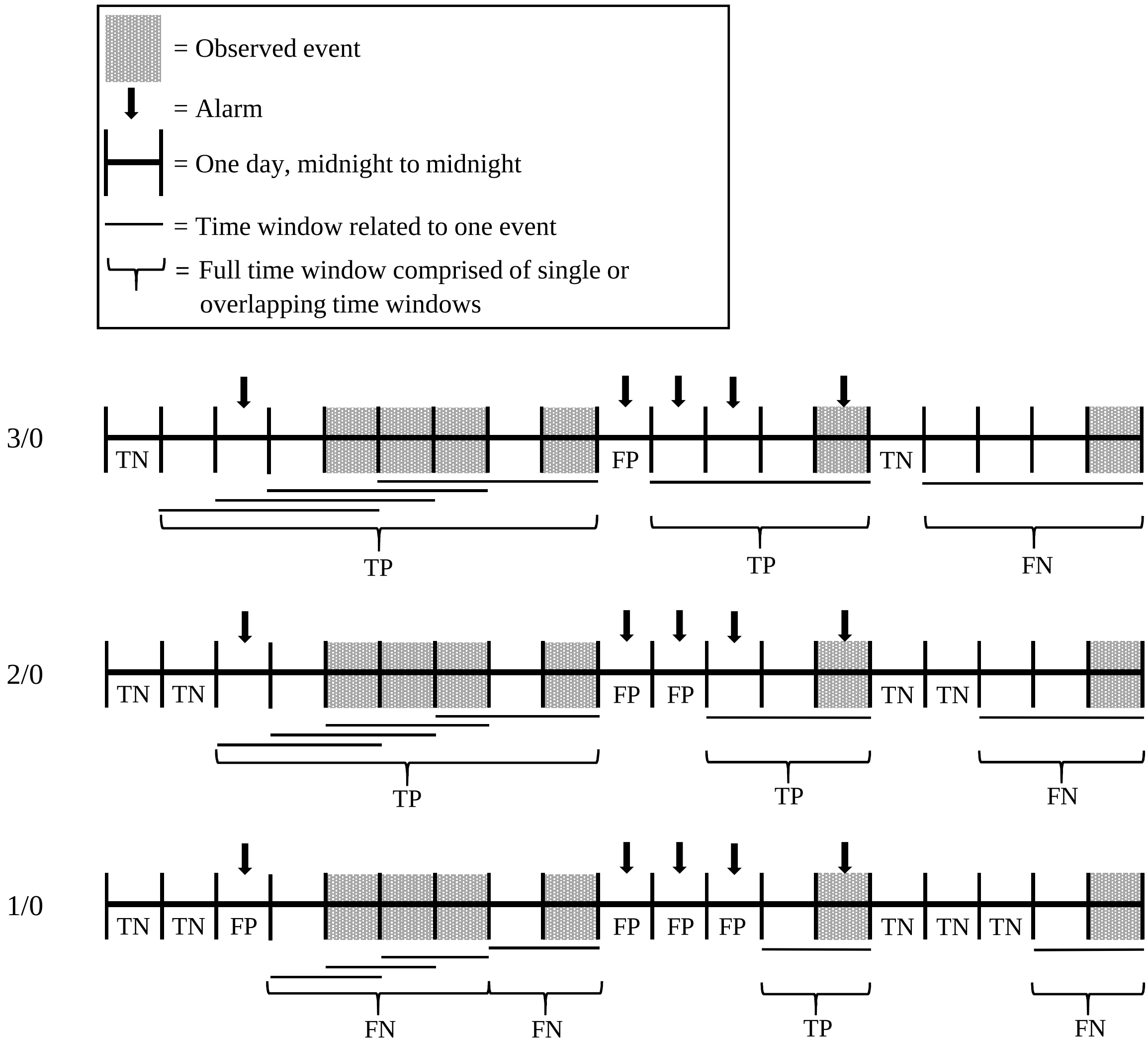


Figure 7.3: Example of definitions of true positives (TP), false positives (FP), true negatives (TN), and false negative (FN). All observed events are associated with a time window, and overlapping time windows are merged into longer windows. Three lengths of time windows are illustrated; 3/0 = three days before an event and zero days after, 2/0 = two days before an event and zero days after, 1/0 = one day before an event and zero days after. All alarms occurring within a time window are counted as one TP alarm. If no alarms occur within a time window, it is counted as one FP. Days outside of time windows but with alarms, are counted as FP, whereas days outside of time windows with no alarms are counted as TN. Based on illustration by Jensen et al. (2017)

7.3.3 *Standardized Two-sided CUSUM*

In a Cusum, the deviations from the mean, μ_0, are accumulated over time, and when the sum of accumulated deviations exceeds a defined threshold, the process is considered out of control and an alarm is generated (Montgomery, 2013).

The inputs to the Cusum in this study are series of forecast errors, e_t generated by the DLM. For a pen it is simply the series of forecast errors from the sensor in the corresponding pen (8 in Herd A, 16 in Herd B), whereas the series of forecast errors for a section (4 in Herd A and in Herd B) is generated by adding the forecast errors of all sensors at time t within the specific section together.

The series of forecast errors for the herd (1 in Herd A and in Herd B) is likewise generated by adding the forecast errors of all sensors in the herd at time t together. In case of missing data at time t, the value of the corresponding forecast error is set equal to zero. Thus, if e_t denotes the full vector of forecast errors at time t, the scalar forecast error e_t^u for the unit u (a specific pen, a specific section or the entire herd) is found as

$$e_t^u = I_u e_t, \tag{26}$$

where I_u is a row vector only consisting of zeros and ones. If u is a specific pen, it means that I_u is a row vector with the element 1 at the position of u in e_t. Accordingly, if u is a section, I_u will have ones at the positions corresponding to pens in the section in question and zeros elsewhere.

The series of forecast variances, Q_t^u, for a given unit are calculated according to standard rules as

$$Q_t^u = I_u Q_t I_u'. \tag{27}$$

In case of missing data at time t, the value of the corresponding forecast variance is sat equal to 1. The cumulated sum of the Cusum is reset when an alarm has been generated, and since the test data for both herds covers the length of two batches, the cumulated sum is also reset at the beginning of the second batch.

In the two-sided Cusum, the forecast errors above the mean (zero) are summed separately as *upper Cusum*, and the forecast errors below the mean are summed separately as *lower Cusum* (see Figure 7.4).

The two-sided Cusum allows for different interpretations of alarms caused by water consumption higher than expected and lower than expected. Since the underlying level of water consumption increases as the pigs grow, the numerical values of the forecast errors increase as well (Madsen et al., 2005). In order to distinguish between the growth-related increase and increases caused by the process being out of control, the forecast errors are standardized, and a *Standardized* two-sided Cusum control chart is applied, as described by Montgomery (2013).

Since the expected value of e_t^u is 0, the standardized value y_t^u simply becomes

$$y_t^u = \frac{e_t^u}{q_t^u}, \tag{28}$$

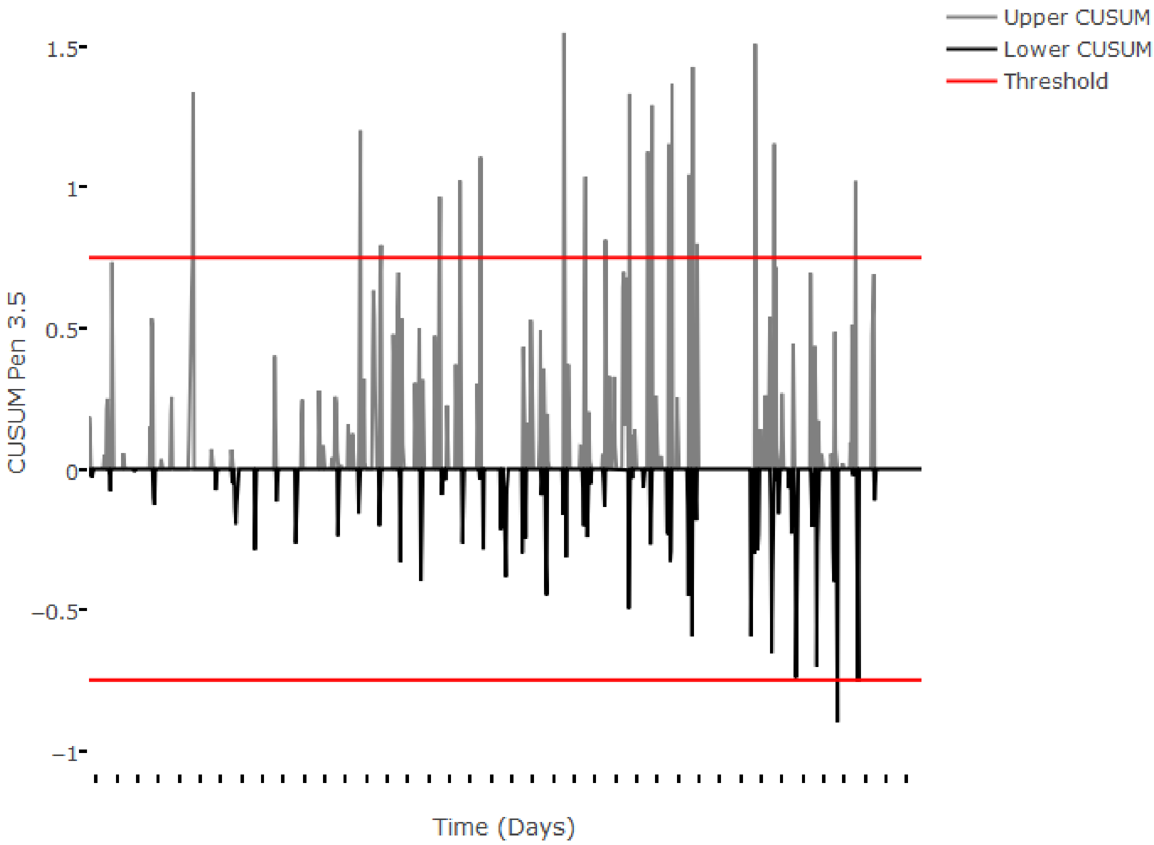


Figure 7.4: Illustration of a two-sided Cusum control chart. The cumulative sum of deviations above the target, $\mu_0 = 0$, is plotted as 'upper CUSUM', and the cumulative sum of deviations below the target is plotted as 'lower CUSUM'. The upper and lower thresholds are defined with equal distance to the target, and the process is out of control when either of the upper or lower Cusum exceeds the threshold.

where $q_t^u = \sqrt{Q_t^u}$.

Then, the **Upper Cusum** for the unit is the series

$$C_t^{u+} = max[0, y_t^u - k + C_{t-1}^{u+}] \tag{29}$$

and the **Lower Cusum** is the series

$$C_t^{u-} = max[0, -k - y_t^u + C_{t-1}^{u-}]. \tag{30}$$

where k is the *reference value*. The reference value allows for a constant level of *slack* or *allowance* to be accepted as an integrated part of the system and it is subtracted from y_t before the summation. The value of k is traditionally chosen relative to the size of the shift to be detected (Montgomery, 2013).

In addition to the reference value, a *decision interval*, or a threshold, h, must be chosen as well. If either C_t^{u+} or C_t^{u-} exceeds the threshold, the process is considered to be out of control, and an alarm is generated. Montgomery (ibid.) recommends h to be defined at fixed values of 4 or 5 for a standardized Cusum. However, when evaluating performances at spatial levels, multiple processes are evaluated, and another approach to defining optimal Cusum parameters is necessary.

Choosing the right settings of the threshold value, h, and the reference value, k, of the Cusum are essential to the number of alarms generated. This is illustrated in Figures 7.5 and 7.6, which show that lower values of either of the Cusum parameters lead to more alarms generated than higher values. Higher values, however, lead to loss of information and failed recognition of the process being out of control. Therefore the optimal combination of h and k for each vector of forecast errors is chosen by iterations over sequences of h and k values. Threshold values were iterated from 0 to 5 and reference values were iterated from 0 to 2.

Evaluating spatial levels

Evaluation of model performance is done for each of the seven model versions (Table 7.2) separately on data from Herd A and Herd B. All model versions are evaluated for their ability to predict the occurrence of events of interest at either of the three spatial levels; pen level (in a specific pen), section level (in a specific section within the herd), or herd level (in any pen within the herd) using three different lengths of time windows.

A pen is defined as the area comprising the number of pigs whose water consumption is monitored by a single sensor (see Table 7.1). A section is defined as all pens with sensors within the same section in the herd. A herd is defined as all pens with sensors within the farm building.

Days with events at pen level are the days when events are registered in the pen by the caretakers. Days with events occurring at section level are all the days with minimum one event registered in any pen within the section. If events are registered in two or more pens within the same section at the same day, they count as one day with a section event. Days with events occurring at herd level

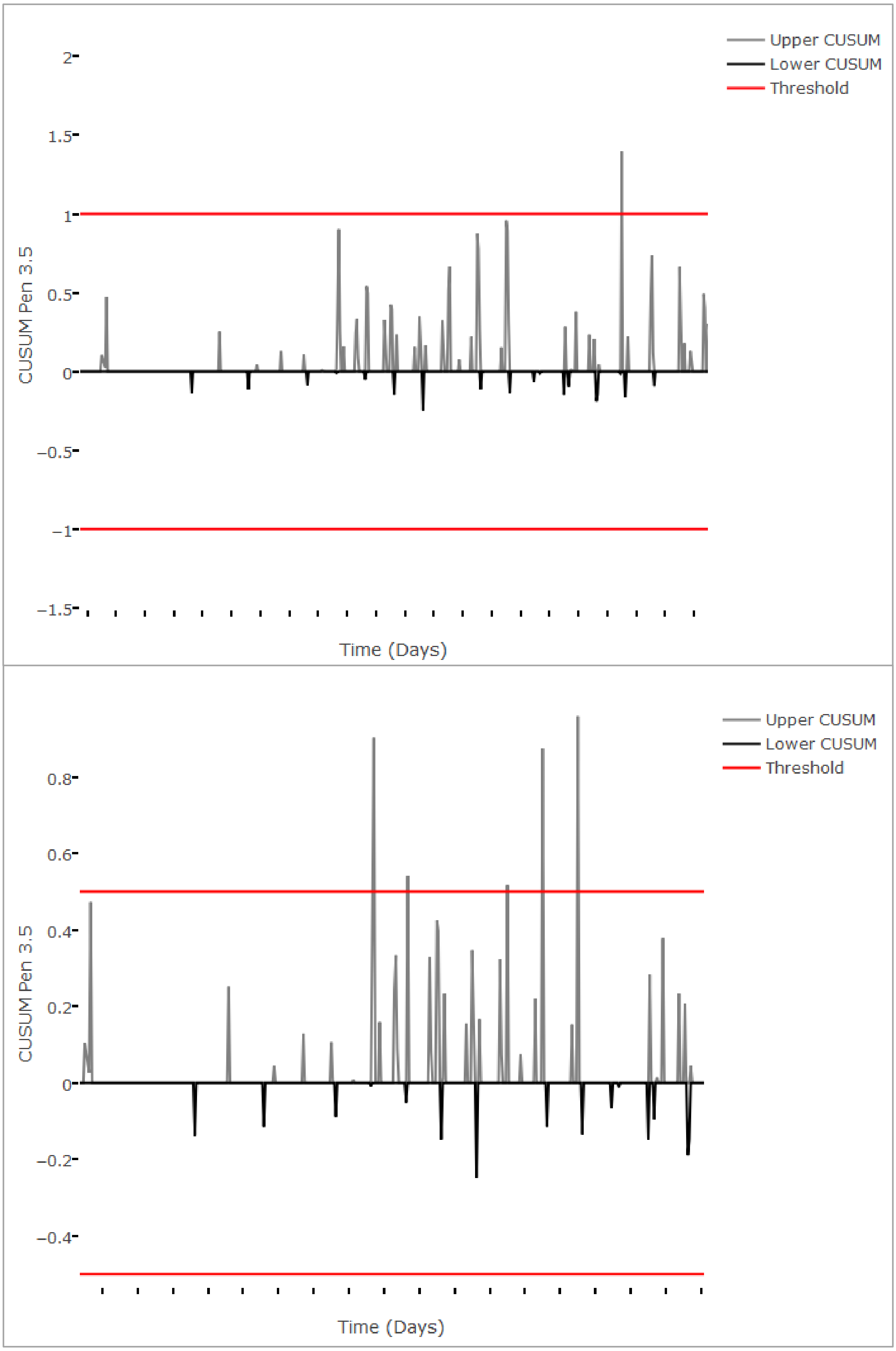
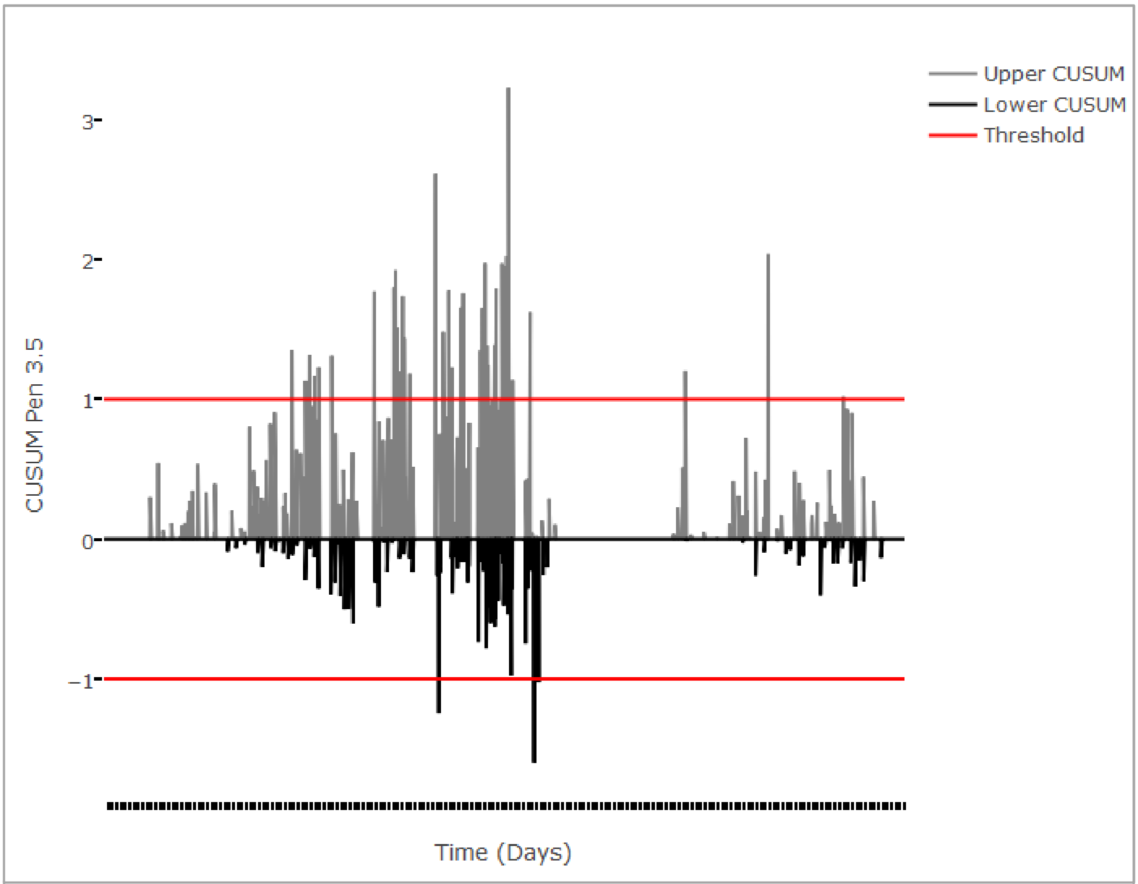


Figure 7.5: Illustration of a two-sided Cusum control chart with different settings of the threshold, h. In the upper plot, $h = 1$, and the Upper Cusum exceeds the threshold one time. In the lower plot, $h = 0.5$, and the Upper Cusum exceeds the threshold five times. The reference value, k, is unaltered in the two plots.



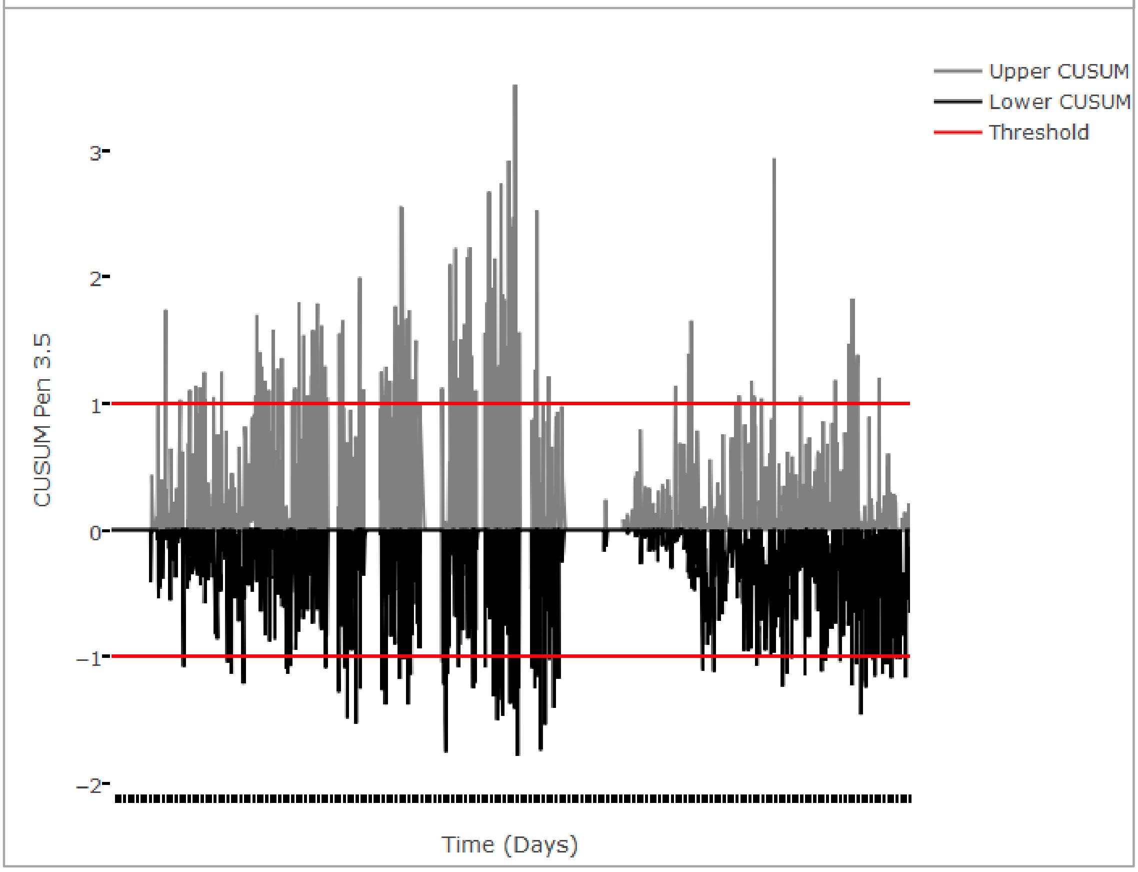


Figure 7.6: Illustration of a two-sided Cusum control chart with different settings of the reference value, k. In the upper plot, $k = 0.5$, and in the lower plot, $k = 0.05$. The reference value is subtracted from each forecast error before they are summed, and more information is lost with the higher k value than with the lower. The threshold, h, is unaltered in the two plots.

are all the days where minimum one event is registered in the herd. If events are registered in two or more pens in the herd at the same day, they count as one day with a herd event.

When evaluating spatial level performance, a total of $2 \times 7 \times 3 \times 3 = 126$ model combinations based on:

- Herd (Herd A, Herd B)
- Model version (HHH, HSP, HSS, SSS, SSP, SPP, PPP)
- Spatial level (Pen, Section, Herd)
- Time Window (3/0, 2/0, 1/0).

A Cusum is run on each series of standardized forecast errors within the level for all of the 126 model combinations per $h \times k$ combination, and the performance is calculated on the pooled outputs of these Cusums per setting. As an example, let n be the number of units (pens or sections) at the spatial level in question and let the *setting*, $s = (h, k)$, be a unique combination of threshold and reference value. Each Cusum with a unique setting, s, is run in turns on standardized forecast errors from all units at the level. In order to obtain the overall performance of the Cusum setting, the four classification categories (TP, FP, TN, FN) are counted across units as follows:

- $TP_s \quad = \quad TP_{s1} + TP_{s2} + \ldots + TP_{sn}$
- $FP_s \quad = \quad FP_{s1} + FP_{s2} + \ldots + FP_{sn}$
- $TN_s \quad = \quad TN_{s1} + TN_{s2} + \ldots + TN_{sn}$
- $FN_s \quad = \quad FN_{s1} + FN_{s2} + \ldots + FN_{sn},$

where e.g. TP_{s1} is the number of true positives for unit 1 under setting s.

Then the conditional prediction accuracy is calculated for each s in terms of sensitivity (Se_s) and specificity (Sp_s) as follows:

$$Se_s = \frac{TP_s}{(TP_s + FN_s)} \tag{31}$$

and

$$Sp_s = \frac{TN_s}{(TN_s + FP_s)} \tag{32}$$

For each spatial level of each model version, the performance parameters Se_s and the false positive rate $FPR_s = 1 - Sp_s$ are plotted against each other so that each setting produces an observation in the diagram as shown in Figure 7.7. For a given value of FPR_s, the best possible Se_s is desired implying that the *Receiver Operation Characteristic* curve (ROC curve) is identified by connecting observations that, for each value of FPR_s, maximize Se_s. Thus, the ROC curve is a nondecreasing function of FPR_s.

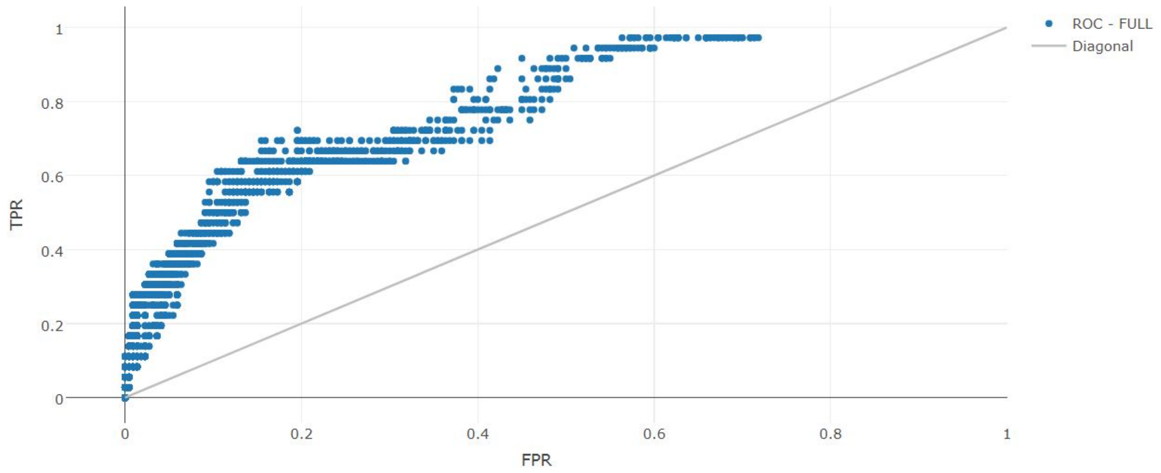


Figure 7.7: Illustration of a ROC curve from a section in Herd B with time window 3/0. The prediction accuracy is plotted for each $h \times k$ combination.

As the final measure of the predictive performance the Area Under Curve (AUC) is calculated. An AUC = 1 indicates perfect predictive performance, so values close to 1 are preferred. The AUC is calculated in R (R Core Team, 2014) using the function "trapz" from the library "pracma".

7.4 RESULTS AND DISCUSSION

The AUC of the 126 different model combinations can be seen in Tables 7.3, 7.4, and 7.5. The predictive performance is in general higher at herd level and decreases as the level gets more detailed, which is illustrated in Figure 7.8 for model version HHH and time window 3/0. The results also show that the AUC in general is higher when the longer time window (3/0) is used, and decreases as the time window gets shorter.

The overall best predictive performances is reached for Herd B, model version HHH at herd level for time windows 3/0 (AUC = 0.9842). For Herd A, the highest predictive performance is reached by model version HHH with time window 3/0 at herd level (AUC = 0.9358).

7.4.1 *Herd level*

Several of the herd level performances (AUC $>$ 0.92) in Table 7.3 indicate prediction accuracy close to perfect. It is, however, worth remembering that any event at any day within the herd is included when evaluating performances at herd level, and that overlapping time windows are merged into one window lasting from the first day of the first window to the final day of the last window. This means that for Herd B, a total of 5 time windows (longest = 47 days) cover 106 of 126 days in the test data when window length 3/0 is used.

For Herd A, a total of 10 time windows (longest = 20 days) cover 79 days out of 172 days in the test data when using window length 3/0. The combination of few time windows and few days outside

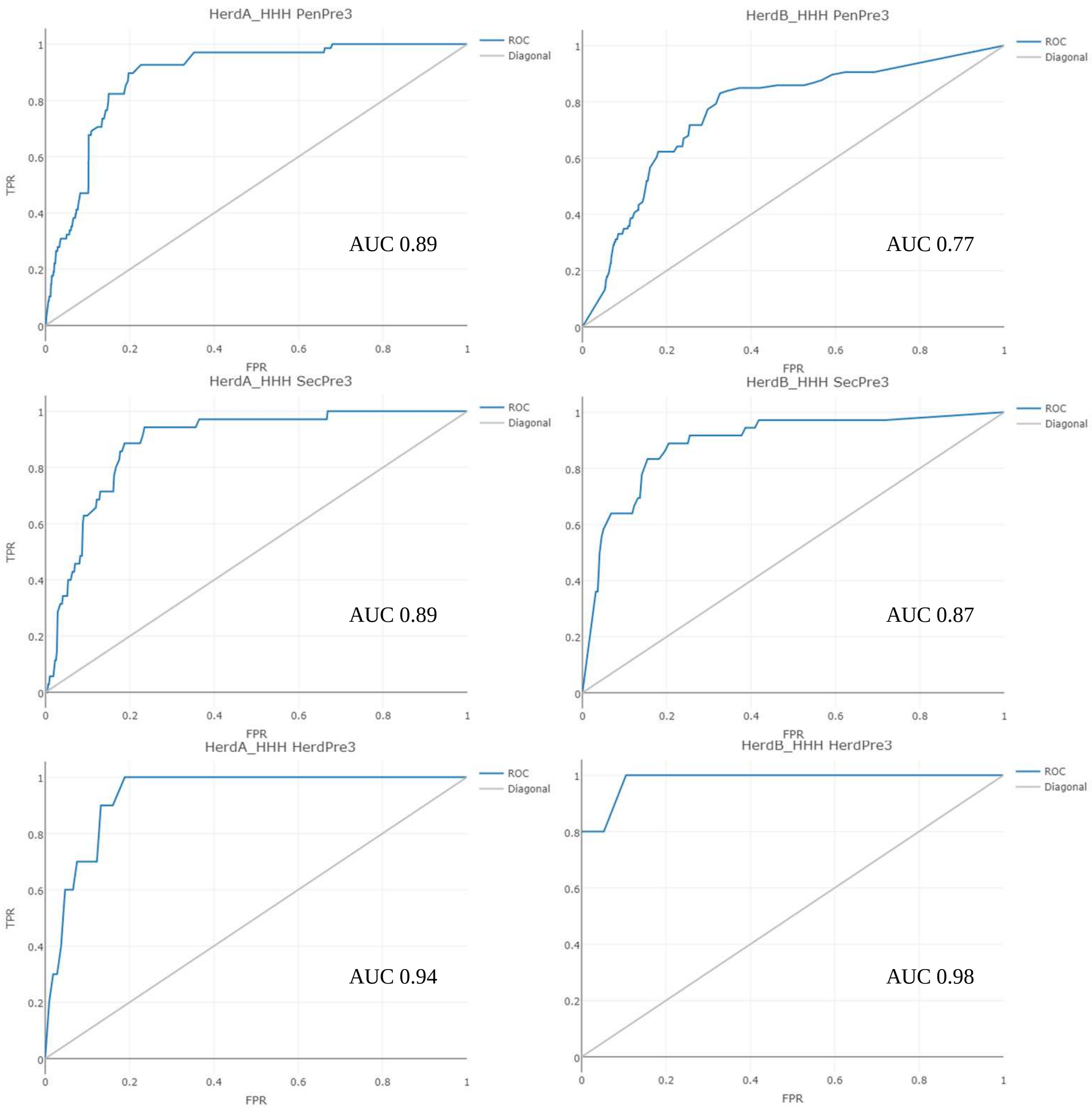


Figure 7.8: ROC curves and corresponding AUC from Herd A (left) and Herd B (right), HHH model version and 3/0 time window. Upper curves: Pen level. Middle curves: Section level. Bottom curves: Herd level.

Table 7.3: AUC (area under curve) for prediction of events at herd level (in any pen in the herd) in Herd A and in Herd B with three different lengths of time windows applied. 3/0 time window covers tree days before the event and zero days after the event, 2/0 time window covers two days before the event and zero days after the event, 1/0 time window covers one day before the event and zero days after the event. AUC for seven model versions is presented for both Herd A (commercial finishers) and Herd B (research centre weaners). Sensors were evenly distributed between four sections in each of the herds with two sensors per section in Herd A and four sensors per section in Herd B. Notations: LG = Linear Growth model, H1 = Cyclic model of length 24, H2 = Cyclic model of length 12, H3 = Cyclic model of length 8. H = Herd level, S = Section level, P = Pen level

Model Structure		Herd A			Herd B		
LG	H1 H2 H3	3/0	2/0	1/0	3/0	2/0	1/0
S	HHH	0.9358	0.9194	0.8013	0.9842	0.9734	0.8878
S	HSS	0.9014	0.8694	0.8287	0.5789	0.6087	0.5369
S	HSP	0.8600	0.8796	0.8307	0.7737	0.7280	0.6761
S	SSS	0.8972	0.8836	0.8083	0.8105	0.8720	0.8438
S	SSP	0.8736	0.8604	0.8052	0.9368	0.9614	0.8473
S	SPP	0.9274	0.9194	0.8036	0.8316	0.8316	0.8253
S	PPP	0.9283	0.9148	0.8090	0.8316	0.8816	0.8395

any time windows affects the outcome of the Se_s and Sp_s leaving only few separate points in the ROC curve. Although the impressive predictive performances are correct, the settings they represent, are of little value considering implementation because an alarm is associated with any event in any pen in the herd within the given time window.

7.4.2 *Section and pen level*

The predictive performances at section and pen level for Herd A are almost identical with respect to model versions and time windows (see Tables 7.4 and 7.5). This indicates a high correlation between pens within the same section regarding both changes in drinking patterns and days with event registration. Event registration data confirm that events in pens within the same section are registered on the same day in Herd A in general. And with the joined water consumption of 36 finisher pigs being monitored per sensor, the drinking pattern is only little affected by changes for one or few pigs.

Fewer and smaller pigs are monitored in Herd B (15 weaners). An irregular drinking pattern from just a single pig in a single pen therefore has larger effect on the water consumption in the pen. Differences between pens are thereby easier generated, and this will be reflected in different AUC's at pen and section level.

Table 7.4: AUC (area under curve) for prediction of events in specific sections in Herd A and in Herd B with three different lengths of time windows applied. 3/0 time window covers tree days before the event and zero days after the event, 2/0 time window covers two days before the event and zero days after the event, 1/0 time window covers one day before the event and zero days after the event. AUC for seven model versions is presented for both Herd A (commercial finishers) and Herd B (research centre weaners). Sensors were evenly distributed between four sections in each of the herds with two sensors per section in Herd A and four sensors per section in Herd B. Notations: LG = Linear Growth model, H1 = Cyclic model of length 24, H2 = Cyclic model of length 12, H3 = Cyclic model of length 8. H = Herd level, S = Section level, P = Pen level

Model Structure		Herd A			Herd B		
LG	H1 H2 H3	3/0	2/0	1/0	3/0	2/0	1/0
S	HHH	0.8882	0.8708	0.8144	0.8715	0.8576	0.7705
S	HSS	0.8592	0.8339	0.8135	0.7193	0.6789	0.6444
S	HSP	0.8616	0.8345	0.8105	0.7711	0.7280	0.6850
S	SSS	0.8647	0.8405	0.8084	0.8205	0.8008	0.7641
S	SSP	0.8611	0.8324	0.8098	0.8635	0.8563	0.8020
S	SPP	0.8757	0.8524	0.7959	0.8375	0.8085	0.7643
S	PPP	0.8631	0.8382	0.7825	0.8311	0.8085	0.7667

Table 7.5: AUC (area under curve) for prediction of events in specific pens in Herd A and in Herd B with three different lengths of time windows applied. 3/0 time window covers tree days before the event and zero days after the event, 2/0 time window covers two days before the event and zero days after the event, 1/0 time window covers one day before the event and zero days after the event. AUC for seven model versions is presented for both Herd A (commercial finishers) and Herd B (research centre weaners). Sensors were evenly distributed between four sections in each of the herds with two sensors per section in Herd A and four sensors per section in Herd B. Notations: LG = Linear Growth model, H1 = Cyclic model of length 24, H2 = Cyclic model of length 12, H3 = Cyclic model of length 8. H = Herd level, S = Section level, P = Pen level

Model Structure		Herd A			Herd B		
LG	H1 H2 H3	3/0	2/0	1/0	3/0	2/0	1/0
S	HHH	0.8878	0.8701	0.8164	0.7671	0.7348	0.6871
S	HSS	0.8599	0.8424	0.8137	0.6468	0.6320	0.6109
S	HSP	0.8598	0.8422	0.8154	0.6747	0.6459	0.6234
S	SSS	0.8583	0.8350	0.8102	0.7535	0.7309	0.7035
S	SSP	0.8585	0.8408	0.8129	0.7682	0.7401	0.6969
S	SPP	0.8782	0.8634	0.8087	0.7750	0.7555	0.7208
S	PPP	0.8644	0.8500	0.7975	0.7671	0.7490	0.7454

In addition, numerous longer periods of missing data throughout the test data set of Herd A may have reduced any differences in drinking patterns between pens, hence promoting similarity between pens and sections. Running the model on data from another herd, or redefining learning and test data in the present data set, is needed to confirm this.

7.4.3 *Time windows*

Longer time windows yield a higher performances for both herds and all model versions. Although an alarm three days before an event (3/0 window) may be too long for the precise timing of managerial interventions, an alarm two days ahead (2/0 window) might be sufficient in many situations - especially if the predictive accuracy is higher than when shorter windows are applied.

When evaluating the 1/0 window performances for Herd A, the HHH model version is able to predict an event in a specific pen with a fairly high predictive accuracy (AUC = 0.8164). Even though the HSP model version predicts events with a higher accuracy (AUC = 0.8307), this is obtained at herd level. As discussed above, the herd level is a very general spatial level at which an alarm constitutes little value in daily management. For Herd B, the highest AUC, given the 1/0 time window, is reached by the HHH model version at all spatial levels. Thus, herd level reaches the highest accuracy (AUC = 0.8878), and then the accuracy is reduced for both section level (AUC = 0.8020), and pen level (AUC = 0.7208).

7.4.4 *Model versions*

The HHH model version provides the highest AUC for predicting events at all spatial levels in Herd A. This indicates that finisher pigs across Herd A show peaks in their drinking pattern at the same time of day (see Table 7.2) throughout the entire growing period. The HHH model version is presenting the poorest fit of the seven versions on Herd A data when developing the model, whereas the SSS model version in general has poor performance in terms of AUC, but obtained the best fit to data (Dominiak et al., 2017a).

No single model version provides the highest AUC across levels in Herd B. For predictions of events at herd level, the HHH model version has the highest accuracy at all time windows, but for predictions at section level and pen level, the model versions with harmonic waves defined at section and pen level provide high accuracies as well.

There is a remarkably clear connection between the level of the harmonic waves in the model versions and the level where events are predicted with the highest accuracy for time window 1/0 in the way that the HHH model version predicts best at herd level (AUC = 0.8878), the SSP model version predicts best at section level (AUC = 0.8020), and the PPP model version predicts best at pen level (AUC = 0.7454).

As for Herd A, it is the HHH model version which has the poorest fit in the DLM and the best prediction accuracy when evaluated on test data and events. The SSP version fit the test data better, but it is the PPP version, which fit the best in Dominiak et al. (ibid.).

An explanation of this inverse relation between fit and prediction accuracy may be that the models with better fit end up overfitting the training data. Over-fitting models tend to model random noise as well, causing the model versions with higher complexity to make worse predictions, when trying to include random noise in the modeled pattern (Fortmann-Roe, 2012).

It may also be, that models with better fit at the same time have a higher adaptability to changes and irregularities. Instead of generating alarms, a well fitting model will adjust to the changes and accept them as a part of the pattern.

7.4.5 *Ensemble classifying methods*

An *ensemble classifier* combines the output of different models and often increase predictive performance over a single model (Witten and Frank, 2005). Alarms communicated from an ensemble are often considered more valid than alarms from an individual model version.

In order to improve predictive performance, the two ensemble classifying methods, *bagging* and *boosting* are tested on the seven model versions from each herd separately. Both methods are machine learning methods, and they combine the decisions of different models by amalgamating the outputs into a single prediction (ibid.). Kamphuis et al. (2010) applied both bagging and boosting to decision trees in a CM detection model. They found bagging to give the better results.

The bagging method lets all model versions vote whether an alarm should be generated or not, on a daily basis. A defined threshold states how many models should agree, and if the threshold is reached or exceeded, the ensemble generates an alarm. The boosting method works on the same principles, only the votes are weighted according to, for example, the performance of each model version. In this test, the specificity of each model was used as weighting factor in the voting under the boosting method.

In our study, neither bagging nor boosting improved the AUC when compared to the AUC of the best of the individual model versions. Thus, the improvement seen in the study by Kamphuis et al. (ibid.) is not seen here. An obvious reason could be that all seven models are based on exactly the same data and, furthermore, have many structural similarities. Thus the seven tests can in no way be seen as independent.

7.4.6 *Alarm prioritizing method*

Some alarms occur at the same time t in pens and in the corresponding section as illustrated in Figure 7.9. If the apparent connection between changes in drinking patterns and general wellbeing (Andersen et al., 2016; Madsen et al., 2005) is accepted, then such alarms should be considered true, independently of event registrations. Such an alarm is either caused by a very large deviation in a

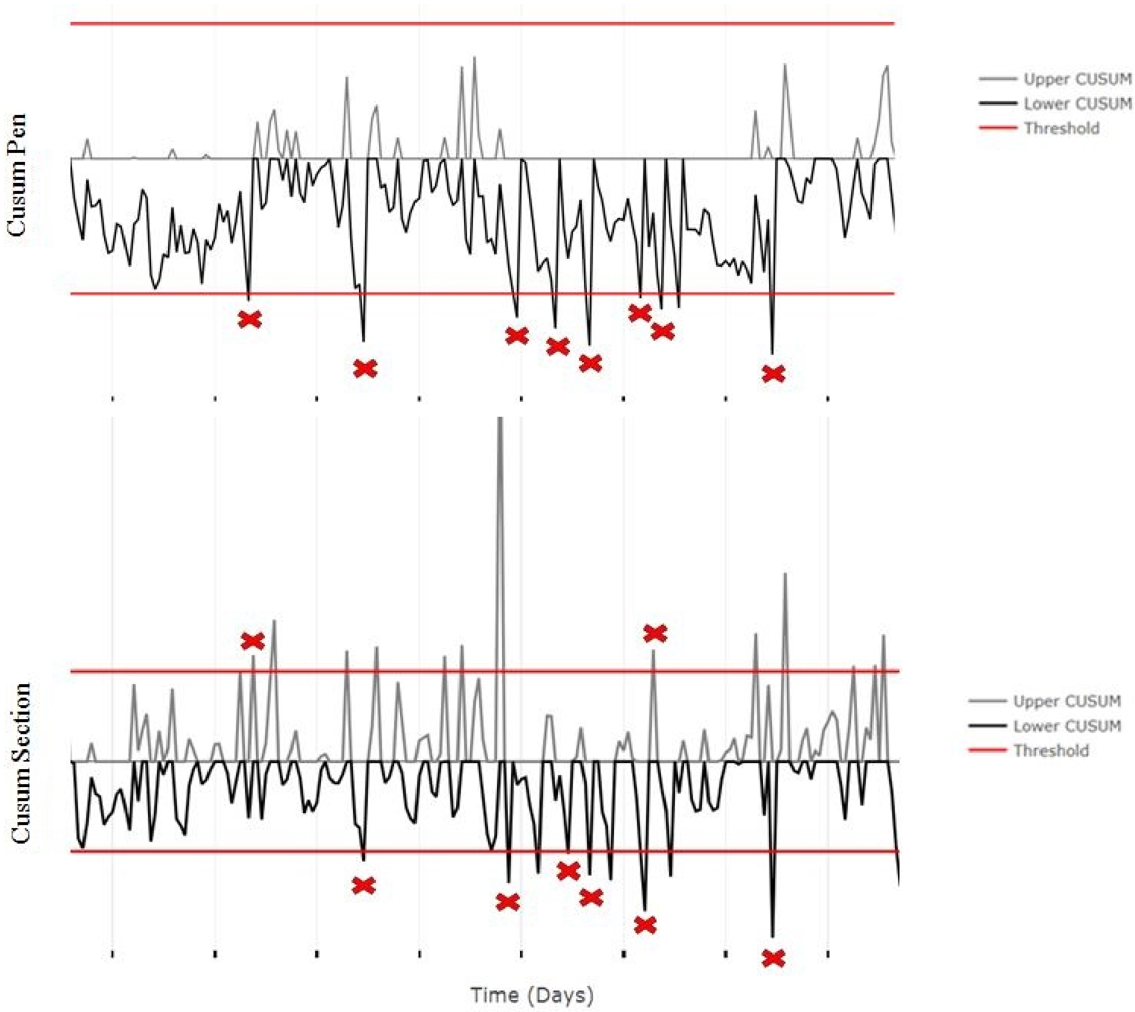


Figure 7.9: Example of a Cusum from a pen (top) and the corresponding section (bottom) with simultaneous alarms. Alarms marked with an X occur at the exact same hour in the pen as in the section.

single pen, or by relatively smaller unidirectional errors in more pens, but they should always be attended.

7.4.7 *Alarm reducing method*

Alarms from multiple pens within the same section on the same day can be merged and communicated as one alarm for the section rather than multiple individual pen level alarms. This method reduces the number of alarms communicated to the manager. Although the method to some extent devaluates pen-specific information, there is a managerial value in section-specific alarms due to the sectionalized structure in the pig producing units as a whole.

7.4.8 *Alternative post processing methods*

DLMs have been used in several previous studies with the purpose of detecting undesired events. The general procedure has been to fit a (univariate or multivariate) DLM to data and, afterwards, to produce series of forecast errors which, in a second step, are post processed in order to produce warnings.

Several different post processing methods have been used previously. Jensen et al. (2017) used a threshold for the Mahalanobis distance (found by Cholesky decomposition of the forecast variance-covariance matrix) between the multivariate forecast error and the zero vector. In another multivariate study, Jensen et al. (2016) used a Naïve Bayesian Classifier and in Jensen and Kristensen (2016), artificial neural networks were used for post processing.

In univariate studies (Cornou et al., 2008; Madsen and Kristensen, 2005) and studies where a multivariate observation has been transformed to a univariate response (Bono et al., 2012, 2013, 2014) a Cusum in combination with a V-mask (Montgomery, 2013) has been a popular tool for detection of gradual changes in the observed pattern of data. For sudden changes a simple Shewhart control chart (ibid.) applied to the forecast errors has often been used (Bono et al., 2012, 2013, 2014; Cornou et al., 2014).

The post processing method used in this study has been the tabular Cusum with various settings but as illustrated by the overview above, many other options exist. It could be argued that using a multivariate approach, and then later only use univariate Cusums for detection of events, considerably reduces the spatial information available in the model. Thus, it would be interesting for future research to study alternative post processing methods.

A first step could be to distinguish alarms generated by the upper Cusum from those generated by the lower Cusum. In case of diarrhea, for instance, an increased water consumption is assumed (Madsen and Kristensen, 2005) and, accordingly, the upper Cusum might generate an alarm. Other disturbances leading to decreased water consumption might produce an alarm generated by the lower Cusum. A more sophisticated approach would be to use a structured Bayesian network for post processing of the forecast error vectors and classify them according to presence or absence of events. That is, however, outside the scope of this article.

7.4.9 *Implementation considerations*

The criterion of sufficient performance, as defined by Hogeveen et al. (2010), is not fulfilled by the present spatial model and evaluation method. The implementation criteria do not consider a spatial approach nor the level of information in drinking patterns on the general wellbeing of pigs (Andersen et al., 2016; Jensen et al., 2017; Madsen et al., 2005). When changes in the general wellbeing is reflected in the forecast errors, alarms are likely to be generated for other causes than the events constituting the gold standard.

Performance evaluation is based on comparing registered events with generated alarms, which may be why no predictive accuracy, or other performance measures, have yet to reach the level defined in the criterion when considering sensor-based detection models (Dominiak and Kristensen, 2017). The spatial approach, described in the present paper and in Dominiak et al. (2017a) aim to generate alarms which are less dependent on a specific event. By focusing on communicating irregularities in a specific spatial area of the herd, the model allows for the manager to combine alarms with knowledge of the animals in the pointed area.

The criterium of similarity between the herds in the study and commercial herds are fulfilled for Herd A since it is a full producing commercial herd. The criterium is fulfilled to some degree for Herd B since 13 full batches from each of 16 pens are monitored, hereby including more animals than is sometimes the case (Dominiak and Kristensen, 2017). In commercial farms, though, around 30 pigs are inserted in a pen of weaners instead of 15 as in Herd B. Had there been twice as many pigs per sensor, less random noise might have been expected.

Based on the length of the time windows none of the model versions with time window 3/0 are very well suited for implementation, when following the implementation criteria. Since overlapping of single time windows often occur when using these longer time windows, extremely long merged time windows can be generated, which are of little value when considering managerial routines.

The 2/0 time window does not fulfill the time window criterium either. But in these authors' opinion, the model's ability to identify a specific pen or section provides significant managerial value using a time window of three days including the day of the event.

Although the 1/0 time window do fulfill the criterium of a 48 hours time window, the prediction accuracy is low. The higher performance and area-specific accuracy associated with the 2/0 time window as compared to the 1/0 time window, makes the 2/0 time window the most implementable of the two.

7.5 CONCLUSION

The spatial approach makes it possible to predict events at separate spatial levels in herds of growing pigs. The model version expressing highest correlation between pens and sections in a herd (HHH) tend to predict better, which may be due to over-fitting of training data in model versions involving lower correlation. Thus, the model providing the best fit to data is not the most well suited for detection of events.

Longer time windows and prediction at herd level yield very high predictive accuracies, but alarms communicated at herd level are of little or no value in a commercial production herd due to very long overlapping time windows and non-specific spatial identification of events.

The predictive accuracies for identifying events in a specific section are high, and combined with a 2/0 time window the multivariate spatial DLM constitute a new and promising approach to sensor based monitoring tools in livestock production.

7.6 ACKNOWLEDGEMENTS

This research was funded by the Danish Council for Strategic Research (The PigIT project, Grant number 11-116191). We further wish to thank the anonymous farmers, the Danish Pig Research Centre and the technical staff (particularly Mads Ravn Jensen) at Aarhus University for installing and supervising the sensors and for taking care of the daily observations in the herd.

CHAPTER 7 REFERENCES

Aarnink, A. J. A. et al. (2006). „Temperature and body weight affect fouling of pig pens“. In: *Journal of Animal Science* 84, pp. 2224–2231. DOI: 10.2527/jas.2005-521.

Andersen, H. M. L., E. Jørgensen, and L. J. Pedersen (2016). „Using Evolutionary Operation technique to evaluate different management initiatives at herd level.“ In: *Livestock Science* 187, pp. 109–113.

Anonymous (2000). *Instruction Leaflet Flow sensors - 15 mm Dia. pipe V8189*. URL: http://docs-asia.electrocomponents.com/webdocs/001b/0900766b8001bb47.pdf.

Bono, C., C. Cornou, and A. R. Kristensen (2012). „Dynamic production monitoring in pig herds I: modeling and monitoring litter size at herd and sow level.“ In: *Livestock Science* 149.3, pp. 289–300.

Bono, C. et al. (2013). „Dynamic production monitoring in pig herds II. Modeling and monitoring farrowing rate at herd level.“ In: *Livestock Science* 155.1, pp. 92–102.

Bono, C. et al. (2014). „Dynamic production monitoring in pig herds III. Modeling and monitoring mortality rate at herd level.“ In: *Livestock Science* 168, pp. 128–138.

Cornou, C., J. Vinther, and A. R. Kristensen (2008). „Automatic detection of oestrus and health disorders using data from electronic sow feeders.“ In: *Livestock Science* 118.3, pp. 262–271.

Cornou, C. et al. (2014). „Dynamic monitoring of reproduction records for dairy cattle.“ In: *Computers and Electronics in Agriculture* 109, pp. 191–194.

Danish Agriculture and Food Council (2010). *Danish Pig Producers and Animal Welfare*. Tech. rep. Danish Agriculture and Food Council.

Dominiak, K. N. and A. R. Kristensen (2017). „Prioritizing alarms from sensor-based detection models in livestock production - a review on model performance and alarm reducing methods.“ In: *Computers and Electronics in Agriculture* 133, pp. 46–67.

Dominiak, K. N., L. J. Pedersen, and A. R. Kristensen (2017a). „Spatial modeling of pigs’ drinking patterns as an alarm reducing method. I. Developing a multivariate dynamic linear model“. In: *Submitted for journal publication*.

Dominiak, K. N., L. J. Pedersen, and A. R. Kristensen (2017b). „Spatial modeling of pigs’ drinking patterns as an alarm reducing method. II. Developing a multivariate dynamic linear model“. In: *Submitted for journal publication*.

Fortmann-Roe, Scott (2012). *Accurately measuring model prediction error*. Essays.

Hogeveen, H. et al. (2010). „Sensors and clinical mastitis - the quest for the perfect alert." In: *Sensors* 10.9, pp. 7991–8009.

Huybrechts, T. et al. (2014). „Early warnings from automatic milk yield monitoring with online synergistic control". In: *Journal of dairy science* 97.6, pp. 3371–3381.

Jensen, D. B. and A. R. Kristensen (2016). „Comparison of strategies for combining dynamic linear models with artificial neural networks for detecting diarrhea in slaughter pigs." In: *Journal of Animal Science* 94.E-Supplement 5, p. 84.

Jensen, D. B., N. Toft, and A. R. Kristensen (2017). „A multivariate dynamic linear model for early warnings of diarrhea and pen fouling in slaughter pigs." In: *Computers and Electronics in Agriculture* 135, pp. 51–62.

Jensen, Dan B., Henk Hogeveen, and Albert De Vries (2016). „Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis". In: *Journal of Dairy Science* 99.9, pp. 7344 –7361. ISSN: 0022-0302.

Kamphuis, C. et al. (2010). „Detection of clinical mastitis with sensor data from automatic milking systems is improved by using decision-tree induction." In: *Journal of dairy science* 93.8, pp. 3616–3627.

Lyderik, K.K. et al. (2016). *Protocol for daily observation in the herds of the PigIT project. PigIT report no 8*. Tech. rep.

Madsen, T. N. and A. R. Kristensen (2005). „A model for monitoring the condition of young pigs by their drinking behaviour." In: *Computers and Electronics in Agriculture* 48.2, pp. 138–154.

Madsen, T. N., S. Andersen, and A. R. Kristensen (2005). „Modelling the drinking patterns of young pigs using a state space model." In: *Computers and Electronics in Agriculture* 48.1, pp. 39–62.

Mein, G. A. and M. D. Rasmussen (2008). „Performance evaluation of systems for automated monitoring of udder health: would the real gold standard please stand up?" In: *Mastitis control: from science to practice. Proceedings of International Conference*. Ed. by T. J. G. M. Lam. The Hague, Netherlands: Mastitis control: from science to practice. Proceedings of International Conference, The Hague, Netherlands; Wageningen Academic Publishers. ISBN: 978-90-8686-085-2.

Mol, R. M. de et al. (1997). „Results of a multivariate approach to automated oestrus and mastitis detection". In: *Livestock Production Science* 48.3, pp. 219–227.

Montgomery, Douglas C. (2013). *Statistical Quality Control A Modern Introduction*. Seventh. John Wiley & Sons, Inc.

Ostersen, T., C. Cornou, and A. R. Kristensen (2010). „Detecting oestrus by monitoring sows' visits to a boar". In: *Computers and Electronics in Agriculture* 74.1, pp. 51–58.

Pedersen, Ken S. (2012). „Smaagrisediarre". In: *Dansk Veterinaer Tidsskrift* 6, pp. 18–22.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `http://www.R-project.org/`.

Rasmussen, M. D. (2002). „Defining acceptable milk quality at time of milking". In:

Witten, I.H. and E. Frank (2005). *Data mining: practical machine learning tools and techniques*. 2nd edition. Morgan Kaufmann, Elsevier.

8

PAPER IV

Accepted for publishing in *Proceedings for the EC-PLF conference in Nantes, 12-14 September 2017*[1]: Dominiak et al. (2017)

CHAPTER 8 REFERENCES

Dominiak, K. N. et al. (2017). „Reducing alarms and prioritising interventions in pig production by simultaneous monitoring of water consumption in multiple pens“. In: *8th European Conference on Precision Livestock Farming*.

[1]**NOTE:** In Table 1, three typos have been corrected in the present version of Paper IV.
These are not corrected in the version printed in the proceedings for the EC-PLF conference in Nantes.
In the caption:
"50 Herd B" is changed to "42 Herd B" and "One section in Herd B" is changed to "One pen in Herd B".
In the table:
In the bottom row of the "Events" column, "7" is changed to "8".

Reducing alarms and prioritising interventions in pig production by simultaneous monitoring of water consumption in multiple pens

K.N. Dominiak[1,2], J. Hindsborg[1], L.J. Pedersen[2] and A.R. Kristensen[1]
[1]*Department of Veterinary and Animal Sciences, University of Copenhagen, Groennegaardsvej 2, DK-1870 Frederiksberg C*
[2]*Department of Animal Science, Aarhus University, Blichers Allé 20, DK-8830 Tjele*
knd@sund.ku.dk

Abstract

Spatial modelling of water consumption in growing pigs can be a useful tool for identifying high risk pens or sections in early detection of diseases and various behavioural problems.

In this study a multivariate dynamic linear model (DLM) is developed based on data from simultaneous monitoring of water consumption across multiple pens in two separate herds. The two herds consist of a commercial finisher herd (Herd A) and a research farm with weaners (Herd B).

Parameters in the model can be defined individually at herd, section or pen level. This spatial distinction allows early warnings to be generated at pen level or merged at section or herd level to reduce the number of alarms. Information on which specific pens or sections are of higher risk of stress or diseases is communicated to the farmer and target work effort to pens at risk.

For Herd A, all model parameters defined at section level resulted in the best fit (MSE = 13.85 litres^2/hour). For Herd B, parameters defined at both pen and section level resulted in the best fit (MSE = 1.47 litres^2/hour).

For both Herd A and Herd B, preliminary results support the spatial approach by generating a reduced number of alarms when comparing section levels to pen levels.

This study is a part of an on-going project aiming to improve welfare and productivity in growing pigs using advanced ICT methods.

Keywords: dynamic linear model, multivariate, spatial, alarm-reducing, drinking pattern, monitoring

Introduction

A variety of sensor based detection models have been designed to monitor production animals and detect specific diseases or conditions (Kamphuis *et al.*, 2010; Garcia *et al.*, 2014; Ostersen *et al.*, 2010). Often the amount of false alarms is too high for the model to be implemented (Hogeveen *et al.*, 2010; Dominiak & Kristensen, 2017), and it has proven to be a difficult yet very important task to reduce the number of alarms communicated to the farmer.

Previous research shows that water withholds important information in prediction of diseases in finisher pigs (Jensen *et al.*, 2017). However, changes in pigs' drinking pattern can also indicate general stress and information on the pigs' wellbeing (Madsen, *et al.*, 2005; Andersen *et al.*, 2016).

For bio-security reasons, Danish pig production units for growing pigs are run with a clear spatial separation between pigs of different age groups. Such a construction of the production site enables a spatial approach where the site is modelled as one production unit (the whole herd) consisting of a number of identical subunits (sections) and each subunit consisting of a number of identical sub-subunits (pens).

The objective of this paper is to present a model which detects unexpected changes in the water consumption of growing pigs across a whole production unit, and produces pen, section- or herd- specific alarms. Simultaneous alarms from pens in the same section, or sections in the herd, are merged which reduces the number of alarms communicated to the farmer.

Material and methods

Data

Data of water consumption (litres/hour) were collected from two herds. Herd A is a Danish commercial finisher herd, and water data from seven batches of pigs were obtained in the period from May 2014 to March 2016 (16309 hours). Herd B consist of the weaner sections of a Danish research facility herd, and water data from 13 batches of pigs were obtained in the period from October 2014 to December 2016 (18755 hours). The sensors were photo-electric flow sensors (RS V8189 15mm Dia. Pipe), and they were placed on the water pipe supplying two neighbouring pens (36 pigs, Herd A) or a single pen (15 pigs, Herd B). In Herd A, eight sensors were placed in two identical pens in each of four identical sections. In Herd B, sixteen sensors were placed in four identical pens in each of four identical sections. In total, eight double-pens from Herd A and 16 single pens from Herd B were monitored during the experimental period.

Every morning, the caretakers at each farm registered events of diarrhoea and fouling, which is a behavioural change where the pigs start to lie on the slatted area of the pen and excrete in the lying area (Aarnink *et al.*, 2006). These event registrations constitute the golden standard together with logbook registrations of unexpected managerial situations affecting the pigs.

General model

The water consumption over time is modelled simultaneously for all sensors in the herd using a multivariate dynamic linear model (DLM) as described by West & Harrison (1999). The observation vector, $Y_t = (Y_{1t}, \dots, Y_{nt})'$, is the amount of water consumed per hour at time *t* for each of the *n* sensors. The relation between Y_t and the underlying parameter vector θ_t at time *t*, as well as the evolution of the system over time, is described through an observation equation and a system equation (Equations (1) and (2), respectively):

$$Y_t = \boldsymbol{F}'_t \theta_t + \nu_t, \quad \nu_t \sim N(\mathbf{0}, \mathbf{V}_t), \tag{1}$$

$$\theta_t = \boldsymbol{G}'_t \theta_{t-1} + \omega_t, \quad \omega_t \sim N(\mathbf{0}, \mathbf{W}_t), \tag{2}$$

The aim of the DLM is to predict the next observation. That is to estimate the parameter vectors, $\theta_1, \dots, \theta_t$, from the observations, $Y_1, \dots, Y_t$. Through every hourly observation of water consumed, the model learns more of the general drinking pattern, and it is constantly updating the amount of information adding the newest observation. Any difference between the predicted observation and the actual observation is withheld in the two error terms, ν_t and ω_t. If the pigs follow their normal drinking pattern and drink as much water as expected, the prediction of the next observation is close to perfect, and any prediction error will be small. If, on the other hand, something is causing the pigs to drink more or less than expected, the prediction error will be larger. A systematic change in the normal drinking pattern will generate a sequence of larger prediction errors, and this will lead to an alarm, which will be described later.

Modelling diurnal patterns

The drinking patterns of both finishers and weaners have clear diurnal characteristics (see Figure 1). Furthermore the underlying level of water consumed increases over time

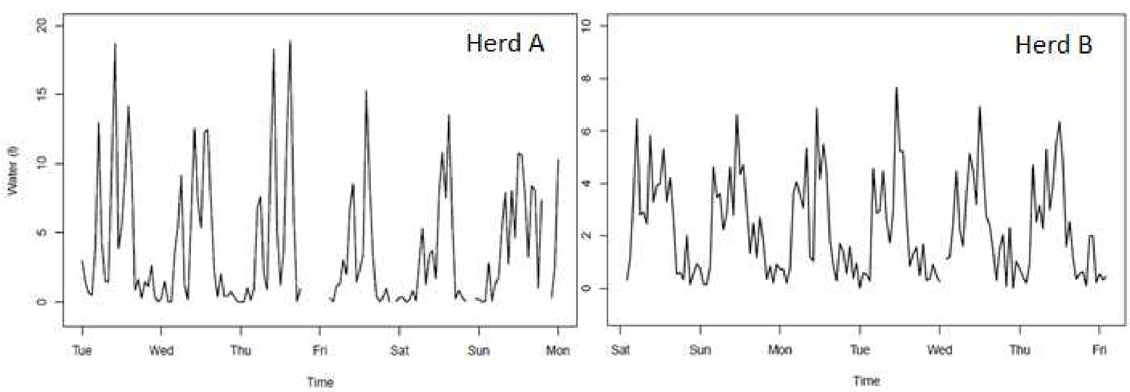


Figure 1: Diurnal drinking pattern of finishers (Herd A) and weaners (Herd B)

indicating that pigs drink more as they grow. A diurnal drinking pattern can be described by the sum of three harmonic waves and an underlying level and trend (Madsen *et al.*, 2005), and the DLM presented here, therefore, consists of four sub-models. The first sub-model, a *linear growth model* (Equation (3)), describes the underlying level and trend,

$$\mathbf{F}_t^l = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{G}_t^l = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{3}$$

whereas the following three sub-models each describes a harmonic wave using the *Fourier form representation of seasonality* (West and Harrison, 1999; Madsen *et al.*, 2005). The Fourier form, as seen in Equation (4), describes a harmonic wave for any frequency, $\omega \in (0, \pi)$, with $\omega = \pi/24$ yielding a wave with a period of 24, $\omega = 2\pi/24$ yielding a wave with a period of 12, and $\omega = 3\pi/24$ yielding a wave with a period of 8.

$$\mathbf{F}_t^h = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{G}_t^h = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix} \tag{4}$$

Modelling spatial structure

Each of the four sub-models can be defined at herd, section or pen level. A sub-model defined at pen level can evolve differently in each pen over time with no interaction between pens. Defined at section level, a sub-model evolves identically in all pens within the same section but differently between sections. Finally, a sub-model defined at herd level evolves identically in all pens in the herd.

The variances components are estimated by the *Nelder-Mead* algorithm in the statistical software R (R Core Team, 2017). The observation variances, $\mathbf{V}_t$, at herd, section and pen level are estimated directly, whereas a system variances, $\mathbf{W}_t$, for each of the four sub-models are estimated through discount factors as described by Madsen *et al.* (2005).

Evaluation

The models are trained on learning data (Herd A: 68 %, Herd B: 83 %) and tested on test data (Herd A: 32 %, Herd B: 17 %) with no pigs delivering data to both data subsets within the herds. Detection of alarms and irregular drinking patterns is done using Tabular CUSUM as described by Montgomery (2013). The standardised cumulated sum (CUSUM) of the positive prediction errors and the negative prediction errors is plotted over time, and if the sum exceeds a defined threshold, an alarm is generated. An event is registered once per 24 hours, but the alarms can be generated at an hourly basis. A '-3/+1' prediction window is defined according to Jensen *et al.* (2017). Hereby all alarms from three days before an event observation to one day after an event observation are merged and considered true positive (TP). If no alarms are generated within the time window, it is considered false negative (FN), whereas single days with alarms but no events are false positive (FP) and single days without alarms but with events are false negative (FN) (see Figure 2).

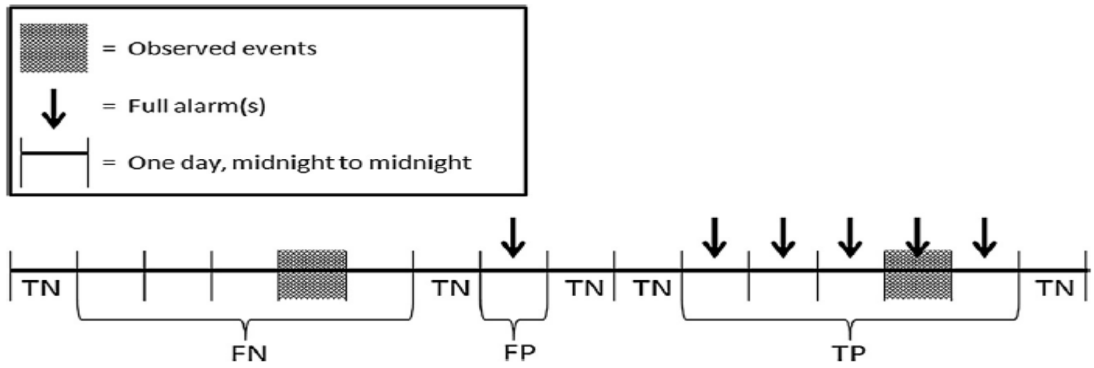


Figure 2: Illustration of the -3/+1 window as described by Jensen *et al.* (2017) with TP,TN, FP and FN alarms defined as described in the text.

Results and Discussion

Model fit

A model with all four sub-models defined at section level, fitted the drinking patterns of pigs in Herd A with MSE = 13.85 litres^2/hour. For Herd B it was a model with the linear growth sub-model defined at section level, and all three cyclic waves defined at pen level, which yielded the lowest MSE (1.47 litres^2/hour). A total of seven models with different level combinations were tested for each herd. In Herd A, each sensor supplies 36 finisher pigs (30-110 kg) which leads to a large variance and hereby a larger MSE. In Herd B, each sensor supplies 15 weaners (7-30 kg), leading to a smaller variance and a smaller numerical MSE.

Detecting events

Based on preliminary results, the spatial DLM is able to detect registered events of either diarrhoea or fouling in both herds. Figure 3 shows how four events were registered in one week in Herd A, and eight in Herd B. Three of the events in Herd A are associated with TP alarms, and all of those would be placed within the same time window, had it been shown. Of the eight events in Herd B, all are associated with TP alarms. No false positive alarms were raised during the week in either herd. The CUSUMS based on prediction errors for a section as compared to prediction errors from individual pens; result in a reduced number of alarms (Table 1). Although the figures presented in this paper are preliminary, there is reason to expect the alarm reducing feature to show in the finished version as well, given the section based production strategy of Danish herds with growing pigs. Because alarms can be generated for the whole herd, a section or a pen, the farmer will be informed of which areas of pigs need extra focus. This can be combined with managerial knowledge of age and health status of pigs in the high risk area. Hereby the right intervention for the given age group of pigs can be chosen.

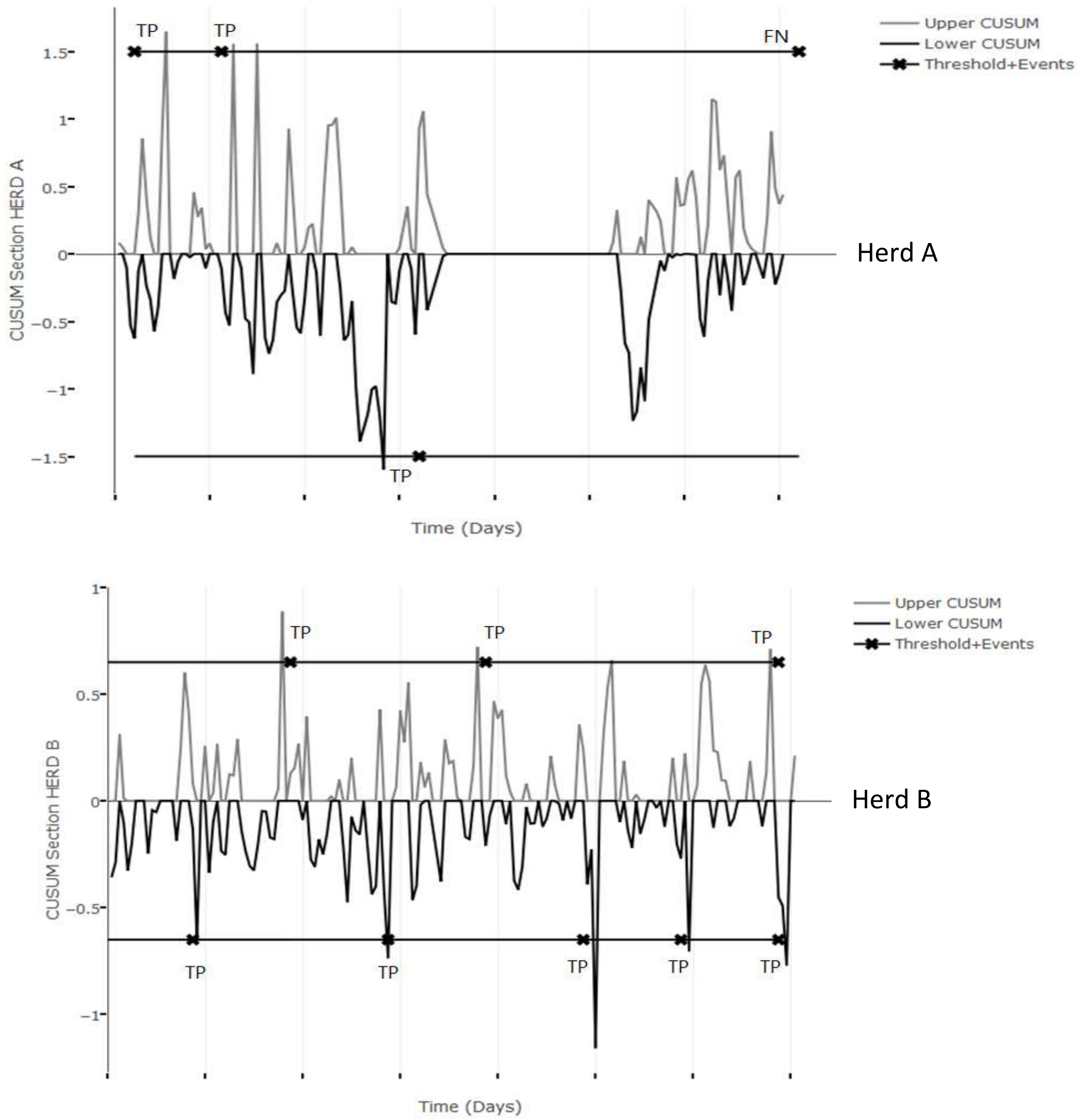


Figure 3: Tabular CUSUM for one week in one section of Herd A and one of Herd B. The two horizontal lines mark the thresholds for the upper CUSUM (grey line) or the lower CUSUM (black line). Four events (marked by **x** on the threshold lines) are registered in Herd A and eight in Herd B. The tabular CUSUM detects three events in Herd A, and eight in Herd B. TP = True Positive, FN = False Negative. The gap around day 5 in the plot is caused by sensor outage.

Table 1: Amount of registered events and CUSUM alarms for one week in pens and the corresponding sections in Herd A and Herd B. With pen level CUSUMS the sum of generated alarms from pens in a section were higher (6 Herd A, 42 Herd B) than with CUSUM at section level (4 Herd A, 8 Herd B). No alarms were merged in time windows; therefore more alarms could be associated to the same event. One pen in Herd B was empty.

Herd	CUSUM level	Events	Alarms	Alarm reduction
A	Pen	3	3	From 6 to 4
A	Pen	3	3	
A	**Section**	**4**	**4**	
B	Pen	7	7	From 42 to 8
B	Pen	7	14	
B	Pen	7	21	
B	**Section**	**8**	**8**	

Conclusion

The preliminary results indicate that a spatial modelling of a pig production herd can reduce the number of alarms communicated to the farmer. Changes in water consumption can be used to identify high risk areas so the farmer can choose the optimal intervention for the pigs in the area triggering the alarm.

Acknowledgements

This research was funded by 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 Centre 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., and Huynh, T.T.T. 2006. Temperature and body weight affect fouling of pig pens. *Journal of Animal Science,* 84:2224-2231

Andersen, H.M.L., Jørgensen, E., and Pedersen, L.J. 2016. Using Evolutionary Operation technique to evaluate different management initiatives at herd level. *Livestock Science,* 187:109-113

Dominiak, K.N. and Kristensen, A.R. 2017. Prioritizing alarms from sensor-based detection models in livestock production – a review on model performance and alarm reducing methods. *Computers and Electronics in Agriculture,* 133:46–67

Garcia, E., Klaas, I., Amigo, J.M., Bro, R., and Enevoldsen, C. 2014. Lameness detection challenges in automated milking systems addressed with partial least squares discriminant analysis. *Journal of dairy science,* 97(12):7476-7486

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

Jensen, D.B., Toft, N., and Kristensen, A.R. 2017. A multivariate dynamic linear model for early warnings of diarrhea and pen fouling in slaughter pigs. *Computers and Electronics in Agriculture,* 135:51-62

Kamphuis, C., Mollenhorst, A., Feelders, A., Pietersma, D., and Hogeveen, H. 2010. Decision-tree induction to detect clinical mastitis with automatic milking. *Computers and Electronics in Agriculture,* 70(1):60-68

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

Montgomery, D. 2013. Statistical Quality Control 7th edition, 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):54-58

R Core Team (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing,* Vienna, Austria.

Sorensen, C.G., Fountas, S., Nash, E., Pesonen, L., Bochtis, D., Pedersen, S.M., Basso, B, and Blackmore, S.B. 2010. Conceptual model of a future farm management information system. *Computers and Electronics in Agriculture,* 72(1):37-47

West, M., and Harrison, J., 1999. Bayesian Forecasting and Dynamic Models 2nd edition, New York, USA: *Springer*.

9

GENERAL DISCUSSION AND PERSPECTIVES

In this Chapter, the overall findings of the thesis are described, and perspectives for future research to improve the presented model, are suggested. Furthermore the challenges of obtaining high model performance in sensor-based detection models will be discussed.

9.0.1 *Findings and perspectives, thesis*

The spatial model, presented in this thesis, does not necessarily reduce the time spent on the daily checking of the animals in a herd. It does, however, point out specific areas in the herd, where the time is better spent at preventive interventions or treatments of the pigs, rather than event-specific alarms for an unspecified area in the herd.

An experienced manager knows of everyday situations, which may affect the animals, and when a high-risk period occurs for each section in the herd. This knowledge can be used to narrow in the plausible causes of an area-specific alarm. However, if the manager is less experienced, or if too many areas are pointed out for various causes, the managerial value of the detection system is weakened.

The risk of event-specific alarms is that the search for clinical signs of the specific event becomes the sole focus of the manager, when attending the alarm. Whereas the strength of area-specific alarms is that a pointed focus-area, allows the manager to concentrate on identifying any signs of impaired health amongst the animals, or signs of suboptimal management regarding for instance feed or climate control.

Although the AUCs of the better performing model versions presented in this thesis, are high ($>$ 0.80), too many false alarms will still be generated. In order to improve the presented detection system, future focus may be targeted at **a)** reducing the number of false alarms by improving model performance, **b)** developing methods for prioritizing or ranking the alarms, and **c)** developing methods for distinction between different events, causing the alarms. More concrete perspectives within these three areas of future focus are suggested in the following.

9.0.2 *Improving model performance*

As discussed in Section 4.2.4, model performance may be improved by a reduction in the complexity of the detection model. However, further actions for improving model performance should not be conducted on the basis of the results obtained in the described studies. Both model fit and detection performance on Herd B data indicated that 15 pigs per pen were too few for the potential of the model to be investigated rightfully. Therefore it would be very interesting to validate the detection system externally on data from one or more independent herds before altering or reducing any parameters in the model. An external validation would test the adaptability of the model and indicate the implemental potential. In addition, external validation would allow the model performance to be evaluated on the basis a gold standard, where events were registered throughout the study period. Both the inconsistent daily registrations in Herd A, and the very low threshold for identifying an event in Herd B, are likely to have affected the model performances in the evaluation of the model.

9.0.3 *Prioritizing alarms*

Different postprocessing methods could be applied in order to differentiate between less urgent alarms and high priority alarms. A prioritizing strategy was described in Paper III and in Section 4.3.4, but methods which combine different sources of data or information would also be interesting to apply. In the work by Steeneveld et al. (2010), sensor based alarms and non-sensor information were combined in order to prioritize the alarms using Bayesian Networks. For the presented model, it would be relevant to investigate the effect of combining area-specific alarms and information on section-specific high-risk periods (See Section 1.2). Such a combination would likely enable the detection system to prioritize alarms associated with high-risk periods higher than other alarms and hereby communicate more precise alarms.

9.0.4 *Distinguishing between events*

When modeling water consumption, a major challenge in distinguishing different causes of alarms is, that water consumption is an indirect indicator of the events of interest. Sensors, which can monitor conditions like diarrhea and fouling directly, do not exist, to this author's knowledge. Instead, the water consumption is monitored, and a working hypothesis states that changes in the drinking patterns of pigs, indicate outbreaks of these specific conditions.

Indirect indicators of unwanted events may, however, be affected by other types of unwanted events than those of interest in the study. Had tail-biting, for instance, occurred in the data, it is likely to have had an impact on the water consumption, as would different causes of lameness or ulcers presumably. However, indirect indicators of events may also be affected by other conditions amongst the pigs, which are not defined as unwanted events. This is assumed to be the case with reduced wellbeing in the presented studies.

The two-sided tabular Cusum may be able to distinguish between events, though, and it is already applied as the postprocessing method for the output of the spatial model. The two-sided tabular Cusum enables a distinction between alarms generated by the upper Cusum and the lower Cusum. Hence, alarms generated by the upper Cusum indicate an increased water consumption, and alarms generated by the lower Cusum indicate a decreased water consumption. Although the findings by Madsen et al. (2005) indicated an increase in the water consumption prior to outbreaks of diarrhea, it is unclear whether a distinction between upper and lower alarms can be used to identify different conditions. The potential of this method to distinguish between different conditions should, however, be investigated.

In addition, further research is needed in order to gain more knowledge on whether reduced well-being of the pigs is reflected as changes in drinking patterns, as indicated by Andersen et al. (2016) and Madsen et al. (2005), and to what extent the drinking patterns are affected. If such an unspecific health impairment, which may or may not progress into a disease, cause a high number of alarms, it is likely that the caretaker will overlook the signs of reduced wellbeing, and misinterpret the alarms as false. This will lead to a devaluation of the information from the detection system.

9.0.5 *Perspectives, performance evaluation*

The research presented in this thesis indicated that the challenge of obtaining high performances from automatic detection models in livestock production may not be possible to fulfill. The literature review, conducted in Paper I, presented an evaluation of sensor-based detection models developed over a twenty-years period. The conclusion was that none of the models were suitable for implementation in commercial herds; the majority due to insufficient detection performances. The spatial detection model, developed in Paper II, did not obtain sufficient levels of performance either, when evaluated in Paper III.

Based on the findings in this thesis, it falls natural to ask the following generic questions:

- Are the minimum performance requirements for the detection of specific events in livestock production realistic to obtain?
- Is the use of performance measurements valid for evaluation of the implemental value of a detection model?

Fully discussing and answering these questions are beyond the scope of this thesis. However, based on the findings and discussions presented in this thesis, three fundamental factors can be identified. These factors point out some challenges, which are likely to impede the performances of detection models in livestock production modeling in general.

The gold standard constitutes the first factor. As discussed throughout the thesis, the imperfect nature of the gold standard constitutes a fundamental challenge for obtaining high model performances. The subjectivity in manual registrations, and the use of a fixed threshold for case vs. non-case definitions of progressive conditions, naturally impede high model performances.

The use of indirect indicators for the event of interest constitutes the second factor. As described above, indirect indicators for unwanted events, may be affected by both other unwanted events and by other conditions, which are not defined as unwanted events. Other indirect indicators, like activity measurements as indicators for oestrus in sows (Cornou et al., 2008; Ostersen et al., 2010), or feeding activity as an indicator for clinical mastitis and lameness (Kramer et al., 2009), are widely used in the models evaluated in Paper I, and they too constitute a risk factor for impeding high model performances.

Biological variation constitutes the third factor, which may impede high model performances. Since the modeling of livestock production imply the modeling of biological variation, some observations will always be extremely high or extremely low given the relevant spectrum. Such extreme observations are random, unpredictable, and unavoidable. Their unpredictability will affect the prediction accuracy to a certain degree, and hereby reduce the performance of the model.

In conclusion, it does not seem realistic to obtain high performances for the detection of specific events based on indirect indicators. Therefore, future research should accept the imperfect conditions of livestock production systems, and focus more on postprocessing and prioritizing methods. The use of sensitivity and specificity as performance measurements for evaluation of the implemental value of a detection model, does not seem valid either. Values of these measurements are always reported given a defined threshold, or warning criterion, and they will therefore always be arbitrary to some extent. Thus, it could be argued that the performance of future detection models should be reported as the unconditional measurement, AUC, instead of specific values of sensitivity and specificity.

CHAPTER 9 REFERENCES

Andersen, H. M. L., E. Jørgensen, and L. J. Pedersen (2016). „Using Evolutionary Operation technique to evaluate different management initiatives at herd level." In: *Livestock Science* 187, pp. 109–113.

Cornou, C., J. Vinther, and A. R. Kristensen (2008). „Automatic detection of oestrus and health disorders using data from electronic sow feeders." In: *Livestock Science* 118.3, pp. 262–271.

Kramer, E. et al. (2009). „Mastitis and lameness detection in dairy cows by application of fuzzy logic." In: *Livestock Science* 125.1, pp. 92–96.

Madsen, T. N., S. Andersen, and A. R. Kristensen (2005). „Modelling the drinking patterns of young pigs using a state space model." In: *Computers and Electronics in Agriculture* 48.1, pp. 39–62.

Ostersen, T., C. Cornou, and A. R. Kristensen (2010). „Detecting oestrus by monitoring sows' visits to a boar". In: *Computers and Electronics in Agriculture* 74.1, pp. 51–58.

Steeneveld, W. et al. (2010). „Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems." In: *Journal of dairy science* 93.6, pp. 2559–2568.

10

CONCLUSION

The overall aim of this PhD project was to investigate whether drinking patterns between pens within a section, and sections within a herd of growing pigs were correlated and could be modeled in a spatial model in order to detect unwanted events in specific areas of a herd.
In Paper I, the literature review presented general difficulties in obtaining high performances for sensor-based detection models in livestock production. None of the detection models were suited for implementation in commercial herds, and the primary reason was found to be insufficient detection performances. The results show that further research is needed on new approaches for reducing alarms from sensor-based detection models in livestock production.
In Paper II, a spatial detection model was developed in order to investigate, whether drinking patterns in different areas of a herd of growing pigs were correlated. The results showed that there was a degree of correlation between the drinking patterns in both the modeled herds, and that the correlation could be modeled. However, the results for Herd B indicated that there were too few pigs in each pen to evaluate the model rightfully, and an external validation of the model would be a first step in identifying how future work on the model should be conducted.
In Paper III, it was found that area-specific alarms could be generated for unwanted events within a short detection window. However, too many false alarms were generated, and it is suggested that future focus on improving the detection system is targeted at **a)** improving model performances, **b)** methods for prioritizing or ranking the alarms, and **c)** methods for distinguishing between different causes of alarms.
In conclusion, the research presented in this PhD thesis, emphasizes the general challenges in obtaining high detection performances for the detection of specific events in livestock production. Especially the use of indirect indicators for the events of interest impedes high performances, as it was also indicated in the presented studies. Based on this PhD project, it is therefore doubtful whether sufficiently high detection performances can be obtained through the modeling of indirect indicators of events of interest alone. A new approach for communicating area-specific alarms as an alternative to event-specific alarms is presented, but further research is still needed in order to investigate the full potential of this method.

LIST OF ABBREVATIONS

AIAO All-In-All-Out

AUC Area Under the Curve

DGLM Dynamic Generalized Linear Model

DLM Dynamic Linear Model

FN False Negative

FP False Positive

ICT Information and Communications Technology

MSE Mean Square Error

NBN Naïve Bayesian Network

ROC Receiver Operating Characteristics

SPF Specific Pathogen Free

TN True Negative

TP True Positive