University of Natural Resources and Life Sciences, Vienna

Department of Sustainable Agricultural Systems

Division of Livestock Sciences

Predicting pregnancy from milk mid-infrared spectroscopy of cow milk

Master Thesis

Philipp Pfeiffer, BSc

Supervisor:

Univ. Prof. DI. Dr. Johann Sölkner

Co-Supervisors:

DI. Dr. Astrid Köck

DI. Lisa Maria Rienesl

Vienna, March 2020

Statutory declaration

I declare that I have prepared, developed and written this thesis independently and I have not used any sources, thoughts or literature of others than clearly stated in the text. The master thesis was not used to award an academic degree at any other university.

_____   _____

Place, date   Signature (Philipp Pfeiffer, BSc)

Abstract

Fertility problems are the most frequent reasons for animal losses in Austrian dairy farms. Pregnancy diagnosis is a very important tool for the reproductive management in efficient and productive dairy farms. Today, the mid infra-red (MIR) spectroscopy is the method of choice in the routine milk recording system for quality control and to analyze the standard milk contents, such as fat, protein, casein, lactose and urea. In some instances, equations based on MIR spectra are used to predict diseases (e. g. mastitis, lameness) and other traits like methane emissions or body energy status. The aim of this study was to develop a calibration equation to predict the pregnancy state from routinely recorded MIR spectral data. A further objective was to evaluate the effect of different sample sizes in calibration. The data for this study was from the Austrian milk recording system and included test day information (e. g. breed, parity, days in milk, milk yield, milk components) as well as information about inseminations and calvings. Test day records of Fleckvieh, Brown Swiss and Holstein Friesian cows between 3 and 305 days of lactation were used for the study. The first derivatives of 212 selected MIR spectral data points, corrected for days in milk by applying Legendre Polynomial, were used for the prediction models. The data set contained about 400,000 records from about 40,000 cows and was randomly split by farm into calibration (50 %) and validation (50 %). The calibration set was balanced (1:1) in terms of pregnant and open cases, while the validation set was kept unbalanced. Pregnancy prediction was done with Partial Least Square Analysis by applying the function 'trainControl'. In a first step, a single prediction equation was applied on all test day records across the whole lactation. The sensitivity and specificity of this model were 0.856 and 0.836. Splitting the results by pregnancy months, showed that sensitivity was very low for the first month of gestation (0.380). Therefore, separate prediction equations for different 'expected' pregnancy stages were created. This strategy led to a much higher sensitivity (0.825) in the first pregnancy months, compared to the sensitivity (0.380) of the model with a single prediction equation. When splitting up the results of this model (separate prediction equation for different pregnancy stages) by lactation month, it was observed that cows which got pregnant very early or very late in lactation were mostly misclassified. Hence, in the final model a separate prediction equation for different 'expected' pregnancy stages and lactation stages was developed. Balanced accuracies were in a range from 0.523 to 0.918. In general, prediction of pregnancy performed better at later pregnancy stages. The evaluation of different sample sizes in calibration demonstrated that sensitivity and specificity was slightly increasing with a higher sample size and standard deviation (sd) was decreasing.

Key words: MIR spectroscopy, pregnancy detection, dairy cow

Zusammenfassung

Fruchtbarkeitsprobleme sind die häufigste Abgangsursache von Kühen in österreichischen Milchviehbetrieben. Eine sichere Fruchtbarkeitsdiagnose ist daher ein sehr wichtiges Instrument eines effizienten und produktiven Milchviehbetriebes. Die Mittlere-Infrarot (MIR) Spektroskopie ist die Methode der Wahl in der routinemäßigen Milchleistungsprüfung zur Qualitätskontrolle sowie zur Bestimmung der Milchinhaltsstoffe wie Fett, Eiweiß, Kasein, Lactose oder Harnstoff. Darüber hinaus werden MIR Spektraldaten auch verwendet, um bestimmte Krankheiten (z. B. Mastitis, Lahmheit, Ketose) oder andere Merkmale, wie Methanemission oder der Körperenergiestatus, vorherzusagen. Das Ziel dieser Masterarbeit war es, eine Kalibrierungsgleichung zu entwickeln, um den Trächtigkeitsstatus von Milchkühen aus den routinemäßig erhobenen MIR Spektraldaten vorherzusagen. Ein weiteres Ziel war die Evaluierung des Effekts unterschiedlicher Stichprobengrößen in der Kalibrierung. Die Daten für diese Studie stammten aus der österreichischen Milchleistungsprüfung und enthielten neben den allgemeinen Testtagsinformationen (Rasse, Herde, Laktation, Laktationstag, Milchmenge, Inhaltsstoffe, ect.) auch Informationen zu Besamungen und Abkalbungen. Für die Studie wurden Testtagsdaten von Kühen der Rassen Fleckvieh, Braunvieh und Holstein verwendet, welche sich zwischen 3 und 305 Laktationstag befanden. Als MIR-Vorhersagevariablen wurden nur selektierte Bereiche des Spektrums verwendet, welche die meiste Information enthalten. Der komplette Datensatz enthielt zirka 400.000 Einträge von ungefähr 40.000 Kühen und wurde nach Betrieb zufällig in einen Kalibrierungs- und einen Validierungsdatensatz geteilt. Zusätzlich wurde der Kalibrierungsdatensatz in Bezug auf den Trächtigkeitsstatus (offen/trächtig) balanciert (1:1). Der Validierungsdatensatz hingegen wurde unbalanciert und somit realistisch belassen. Die Vorhersage erfolgte mit der Methode Partial Least Square Analysis (PLS) unter Anwendung der Funktion 'trainControl' vom R package ‚caret'. Im ersten Schritt wurde eine Vorhersagegleichung für alle Kühe und deren Testtagsergebnisse über die gesamte Laktation entwickelt. Die Sensitivität und Spezifizität bei diesem Modell lag bei 0,856 bzw. 0,836. Die Aufspaltung der Ergebnisse in einzelne Trächtigkeitsmonate zeigte jedoch, dass die Sensitivität im ersten Trächtigkeitsmonat nur bei 0,380 lag und mit fortschreitender Trächtigkeitsdauer anstieg. In einem nächsten Schritt wurde daher eine separate Vorhersagegleichung für jedes (erwartete) Trächtigkeitsmonat erstellt. Bei diesem Modell stieg die Sensitivität im ersten Trächtigkeitsmonat von 0,380 auf 0,825 an. Eine genauere Betrachtung der Ergebnisse zeigte aber, dass Kühe welche sehr früh bzw. sehr spät in der Laktation trächtig wurden, sehr häufig falsch eingestuft worden sind. Um diesen Effekt auszugleichen, wurde im finalen Modell eine separate Vorhersagegleichung für jedes (erwartete) Trächtigkeitsstadium sowie jedes Laktationsstadium entwickelt. Die Ergebnisse (Balanced Accuracies) dieses Models lagen breit gestreut im Bereich zwischen 0,523 und 0,918. Im Allgemeinen

war die Vorhersage in späteren Trächtigkeitsstadien genauer, als jene in früheren Trächtigkeitsstadien. Die Evaluierung des Effekts unterschiedlicher Stichprobengrößen bei der Kalibrierung zeigte, dass Sensitivität und Spezifität mit einer steigenden Stichprobengröße leicht zunahmen und die Standardabweichung abnahm.

Schlüsselwörter: MIR Spektroskopie, Trächtigkeitsvorhersage, Milchkuh

Acknowledgment

Table of content

List of tables

List of figures

List of abbreviations

AUC	area under the receiver operating characteristic curve
AI	artificial insemination
bal.acc.	balanced accuracy
C	calving
DIM	days in milk
ELISA	enzyme-linked immunosorbent assay
FPR	false positive rate
GL	gestation length
MIR	mid-infrared
O	open
P	pregnant
PAGs	Pregnancy-associated glycoproteins
PLS	partial least square analysis
ROC	receiver operating characteristic
SCC	somatic cell count
sd	standard deviation
sens.	sensitivity
spec.	specificity
test	validation set
TPR	true positive rate
training	calibration set

1. Introduction

1.1. General background

This study is part of the international project D4Dairy. This project is led by the Rinderzucht Austria with the overall goal to provide digital support to dairy management via digitalization, data integration, detection and decision support to further improve animal health, nutrition, animal welfare and product quality. The D4Dairy subprojects are divided into two areas. The goal of the first area is Data and Decision. In this area, digitalization, data integration and decision support are the main parts. The second area is Data and Detection. The goal of this second part is to improve health by data-driven detection of risk factors and early predictors. The sub goal 2.2 is defined as the use of milk mid-infrared (MIR) spectroscopy to predict the health of dairy cows. In this sub goal the pregnancy detection has been included additionally (D4Dairy, 2020).

Currently, the MIR spectra is the method of choice in the milk recording system to analyze the standard milk components including fat, lactose, protein and urea (Grelet et al. 2015, 2016). Moreover, MIR spectra data could also be used to predict some other detailed milk components such as lactoferrin (Soyeurt et al. 2012), minerals (Soyeurt et al. 2009) or fatty acids (Soyeurt et al. 2011). It is well known that there are changes in milk yield and also milk components during the pregnancy in dairy cows (Olori et al. 1997). Therefore, the focus of this study was the prediction of pregnancy stage with MIR spectra data, because pregnancy and pregnancy diagnosis are among the most important elements in a successful reproduction management on dairy farms (e.g. Balhara et al. 2013, Pohler et al. 2016, Hirpa et al. 2018). The MIR spectra pregnancy probabilities could be an extra information for the farmers in the framework of the routine milk recording system. So this would allow the farmers to be informed about the pregnancy state of individual cows on each test day and therefore the reproduction management could be optimized.

1.2. Aim of the thesis

The main aim of this study was to develop a calibration equation to predict the pregnancy state from routinely recorded MIR spectral data and to evaluate its specificity and sensitivity. Further, we aimed to provide a formula to give the probabilities of pregnancy for each test day. Further, we evaluate the effect of different sample sizes in the calibration setting.

1.3. Literature review

1.3.1. Pregnancy Diagnosis in Dairy Cows

Pregnancy diagnosis is a very important tool for the reproductive management in efficient and productive cattle farms. On an economical dairy operation, cows need to calve every year and for this reason the identification of pregnant and non-pregnant animals at an early date is necessary (Hirpa et al. 2018). Furthermore, fertility is the most frequent reason for animal losses in dairy farms. According to Egger-Danner et al. (2018), fertility was with an average of 24.2 % the most common cause for dairy cows leaving the herd. For these reasons, an ideal pregnancy test should have high sensitivity, which is defined as the proportion of pregnant cows predicted as pregnant, as well as specificity, which is defined as the proportion of non-pregnant cows predicted as non-pregnant, and should be inexpensive and simple carried out under field conditions (Pohler et al. 2016). Basically, there are two types of pregnancy diagnosis methods available:

Direct Methods

1. <u>*Estrus detection:*</u> This is a simple method and in cattle the estrus is generally known as heat. If the insemination was not successful, the cow does return to estrus 18 to 24 days after breeding. A farmer does not need specialized skills or instruments but there are many factors that limit its accuracy like lactation state or undernutrition. For a good pregnancy detection, it is very important to be properly trained to recognize heat behavior (Pohler et al. 2016).

2. <u>*Transrectal palpation:*</u> This direct method is very common in dairy cows and it was first described by Cowie (1948). Traditionally, the examination is not earlier than 40 to 60 days after the insemination but a good trained examiner can predict pregnancy as early as day 30. It also depends on the size and the age of the dam. When a cow is pregnant, some changes in the location, content, size and texture of the uterus occur and the examiner can declare the pregnancy state of the cow. It is also possible to estimate the embryonic or fetal age and to detect the ovarian structures (e.g. corpus luteum or follicle). There are various published evidences regarding to the effect of the examination on the embryo or fetus but in conclusion, the risk of embryo or fetal loss is less than the value of the gained information (Fricke et al. 2016, Pohler et al. 2016).

3. <u>*Transrectal ultrasonography:*</u> This third direct method has made pregnancy diagnosis possible as early as day 25 after insemination. The fetal heartbeat can be detected as of day 21 but it is easier and safer after day 25. With the use of ultrasonography, the pregnancy state, fetal sex, ovarian function and uterine morphology can be evaluated. The advantages of this method

are also that it is less invasive than transrectal palpation and the pregnancy detection can be made in an earlier pregnancy state (Pohler et al. 2016).

Indirect Methods

4. *Progesterone:* Progesterone is an important steroid hormone and it is involved in the female reproduction cycle, pregnancy and embryogenesis. At the beginning of the pregnancy it is primarily produced and secreted by the corpus luteum and from about day 70 the placenta produces significant quantities (DeWitt & Grondin 2011). The concentrations of progesterone in blood or milk vary with the estrous cycle. The maximum value is 14-15 days after estrus and if the cow is pregnant, this high level continues. So low progesterone levels at 18 to 24 days after insemination can indicate non-pregnant cows. In Figure 1 the representative progesterone level of an open cow (orange line) and a pregnant cow (blue line) is shown. Based on these lines it is not accurate to use the progesterone level to detect pregnant cows because of the high level during the normal estrous cycle (Balhara et al. 2013, Fricke et al. 2016, Pohler et al. 2016). For these reasons of inconsistences, the pregnancy diagnosis based on progesterone has not been heavily adopted in the agricultural practice (Pohler et al. 2016).
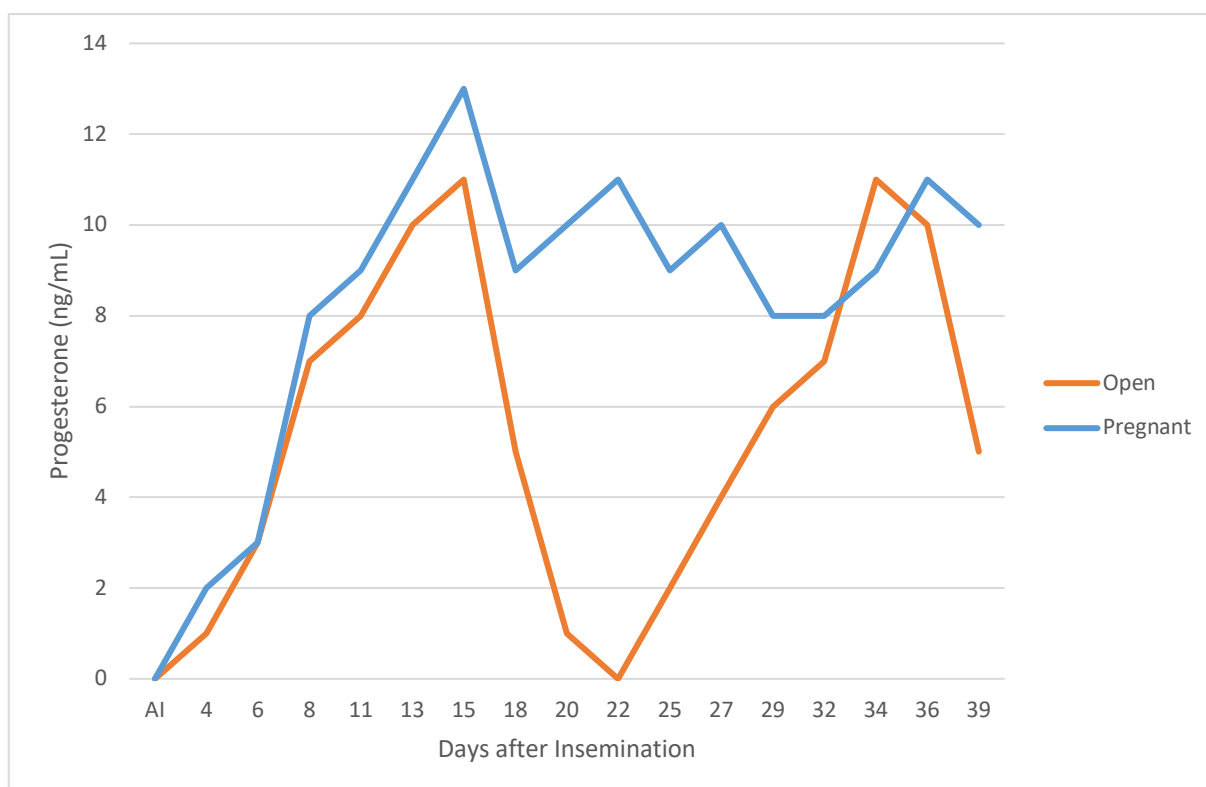
Figure 1: Representative progesterone profiles from blood samples (according to Fricke et al. 2016)

5.   *Pregnancy-associated glycoproteins (PAGs):* PAGs are produced by the binucleate trophoblast cells of the placenta and entered to the maternal circulation (plasma and milk) at day 22 to 24 after the successful insemination. To get an acceptable accurate pregnancy diagnosis, this test is made earliest at day 28 (Pohler et al. 2016). To analyze the PAG value of milk and plasma, a commercial enzyme-linked immunosorbent assay ELISA kit is used. Commercial ELISA kits are offered by different companies. One example for a commercial PAG ELISA manufacturer is the IDEXX milk or plasma pregnancy test (IDEXX Laboratories). The results of this test are separated into three different classes. In plasma, a cow is open when the results are <0.300 and the cow is pregnant when the result is >1.000. Between 0.300 and 1.000 the cow is classified as "recheck". In milk, a cow is classified as open when the results are <0.100 and the cow is classified as pregnant when the results are >0.250. Between 0.100 and 0.250 the cow is classified as "recheck" (Ricci et al. 2015). In Table 1 there are the sensitivity and the specificity obtained using PAG tests based on several experiments. In Austria, the IDEXX milk pregnancy test is very common. This test is also offered by the Austrian milk recording system.

Table 1: Sensitivity and Specificity for ELISA PAG test results based on several studies

Reference	Days after AI	Test	Sensitivity [% (no./no.)]	Specificity [% (no./no.)]
Lawson et al. 2014	33-52	ELISA (milk PAG)	100 (65/65)	98 (46/47)
Sinedino et al. 2014	28-30	ELISA (plasma PAG)	96.1 (173/180)	90.7 (204/225)
Ricci et al. 2015	32	ELISA (plasma PAG)	100 (57/57)	87 (73/84)
		ELISA (milk PAG)	98 (52/53)	83 (68/82)

6.   According to Balhara et al. (2013) there are several other indirect indicators for pregnancy but they are not frequently used in practice and therefore they are only listed and not described further:

1.   Estrone Sulphate
2.   Conceptus and Placenta Secreted Products
3.   Early Conception Factor
4.   Interferon-Tau

1.3.2. Pregnancy Losses

As already mentioned, fertility is a very important part for a successful dairy farm and pregnancy losses play an important role. Generally, the gestation is divided into two periods. The embryonic period starts with the conception and ends with the differentiation stage, which is approximately at day 42-45 of gestation. Then the fetal period starts and this period ends with the birth of the calf (Committee on Bovine Reproductive Nomenclature, 1972). For this reason, pregnancy losses are also divided into these two periods. According to a review paper of 14 studies, the embryonic mortality in dairy cows averaged 12.8 % based on transrectal ultrasonography (Santos et al. 2004). Another study by Humblot (2001) averaged the early and the late embryonic pregnancy losses in Holstein cows in 44 herds in France after first insemination to 31.6 % and 14.7%. The average of late embryonic and fetal losses and their ranges of 17 studies are shown in Table 2. According to another study, fetal losses from day 56 to day 98 of gestation were 2 % (Vasconcelos et al. 1997).

Table 2: Late embryonic and fetal losses in lactating and primigravid dairy and beef cattle (according to Santos et al. 2004)

	Dairy cattle	Beef cattle
Lactating cow	10.70 (8.3-24.0)	9.17 (6.5-10.8)
Primigravid	2.52 (1.5-10.2)	4.21 (4.0-4.8)

These pregnancy losses reduce the benefit of the early pregnancy diagnosis. One or more subsequent pregnancy diagnoses may be necessary to identify these pregnancy losses and to re-inseminate the cows. Therefore, many dairy farmers do a pregnancy diagnosis around 28 to 35 days after insemination and then a second diagnosis around 4 to 6 weeks later to confirm the result (Fricke et al. 2016). Therefore, the prediction of the pregnancy state of a cow every test day from analysis of MIR spectra would be a truly useful tool for farmers.

1.3.3. Milk mid-infrared (MIR) spectroscopy

The need for secure, fast, cheap and high-throughput analyzation methods has led to the application of infrared spectroscopy in the livestock and also in the food sector (De Marchi et al. 2014). Today, the MIR spectroscopy is the method of choice in the routine milk recording system, for quality control and to analyze the standard milk contents, such as protein, casein, fat, lactose and urea (Grelet et al. 2015, 2016). With MIR it is also possible to analyze some other milk traits like minerals (Soyeurt et al. 2009), fatty acids (Soyeurt et al. 2011) or lactoferrin (Soyeurt et al. 2012). The technique is based on the interaction between electromagnetic waves and matter. The different spectral regions can be seen in

Figure 2. The x-ray region is in the wavelength region 0.5 to 10 nm, the UV region in 10 to 350 nm, visible region in 350 to 800 nm, near-infrared region in 800 to 2,500 nm, mid-infrared region in 2.5 to 25 µm, microwaves in 100 µm to 1 cm and radio frequency regions in 1 cm to 1 m (De Marchi et al. 2014).
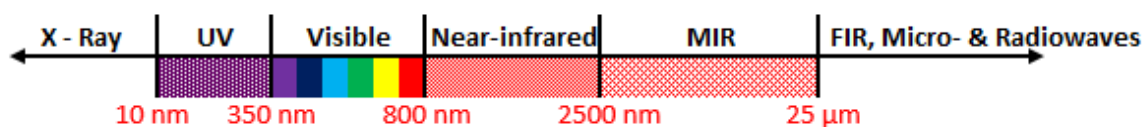
Figure 2: Different spectral regions according to De Marchi et al. 2014

According to the international norm ISO 9622:2013 for milk and liquid milk products, the milk sample is analyzed after pretreatment and homogenization in a so-called infrared spectrometer. That instrument records the quantity of radiation in transmittance at specific wavelengths in the MIR region. These obtained spectral data are transformed into estimates of constituent concentrations or other physico-chemical parameters through calibration models (ISO, 2013). Figure 3 shows a typical MIR absorption curve of a milk sample.
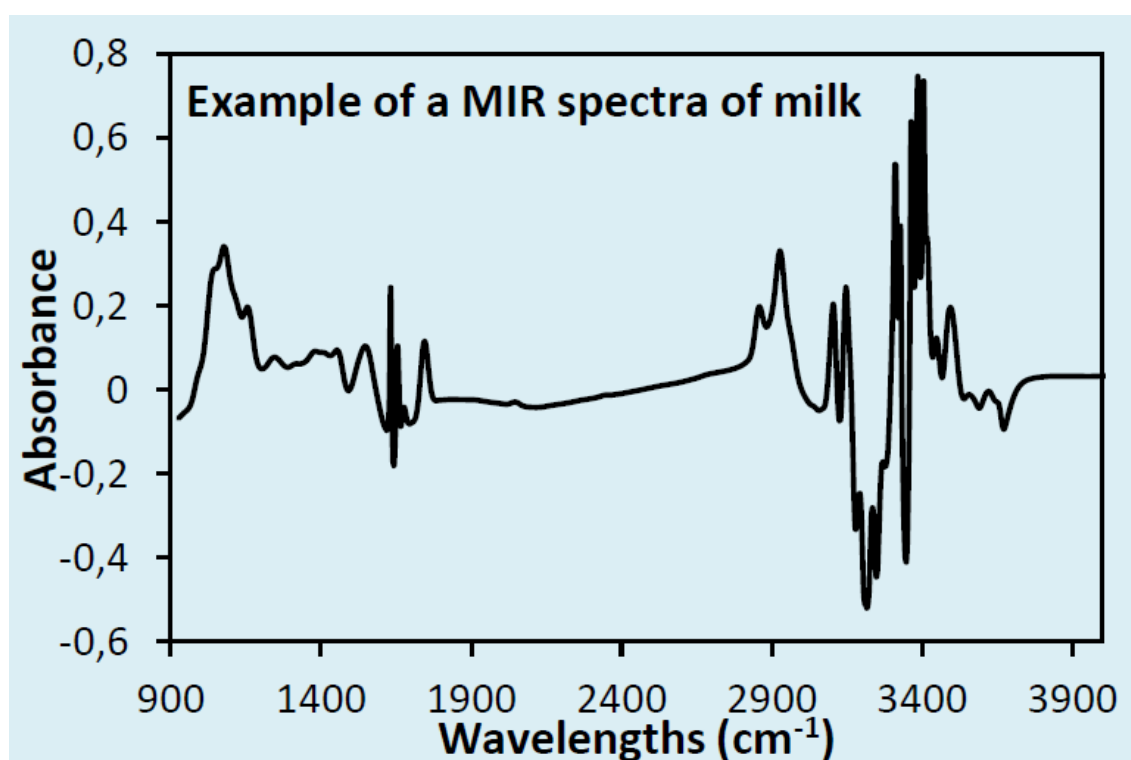
Figure 3: Typical milk MIR absorption curve (Source: OptiMIR)

As already mentioned, MIR spectroscopy is not only used to analyze the standard milk components or other milk traits. Because of the increasing importance of fitness and health traits in national breeding programs, there are several studies to predict diseases with MIR spectroscopy, e.g. mastitis (Dale & Werner 2017, Rienesl et al. 2019), lameness (Mineur et al. 2017, Bonfatti et al. 2019) or ketone bodies

(De Roos et al. 2007). Further, there are studies on methane emissions (Vanlierde et al. 2015), body energy status (McParland et al. 2011), energy intake and efficiency (McParland et al. 2014) or ration composition (Klaffenböck et al. 2017).

There are only a few studies available on prediction pregnancy with MIR spectra. According to a study from Laine et al. (2013), changes in the MIR spectra can provide indications of the change in the pregnancy state of a dairy cow. The aim of this study was the development of a formula to identify open cows from pregnant after an insemination within 20 to 50 days by using the MIR spectra data. Residual spectra were used, defined as observed spectra minus expected spectra. The expected spectra were calculated before in a mixed model. The sensitivity and the specificity are given in Table 3.

Table 3: Sensitivity and Specificity of the pregnancy prediction with MIR spectral data according to Laine et al. (2013)

Days after insemination	Sensitivity [%]	Specificity [%]
21 – 30	99.2	89.1
31 – 40	99.5	83.6
41 – 50	99.7	84.2

Another study from Toledo-Alvarado et al. (2018) used the whole raw MIR spectrum or some other milk components such as fat, protein, casein or lactose to diagnose the pregnancy state. In this study, only test day records after an insemination and with a calving event were used for the model. They concluded that the prediction of pregnancy by using the MIR spectrum is difficult because of the indirect correlation of the pregnancy state with the milk compositions. The results [area under the receiver operating characteristic curve (AUC)] of this study were in a range of 0.607 in Holstein and 0.645 in Alpine Grey.

According to another study from Delhez et al. (2020), the pregnancy prediction with MIR spectra gives promising accuracy in the late stage of pregnancy but not in the early one. In this study, three strategies were applied. In the first strategy, they only used test day records after an insemination. In comparison to Toledo-Alvarado et al. (2018), they also used cows with no calving event after insemination. The AUC value in the test set was 0.650. They concluded this poor performance of the model, that the spectra after an insemination was too noisy and contained not only the effect of pregnancy, such as genetics, management or lactation stage. In the second part, they used the spectral differences (residual spectra), which were defined as a spectrum before minus a spectrum after an insemination. The aim of this strategy was to eliminate the noisy effects. The AUC value, using the residual spectra (0.580) were not observably different from those which used only spectra after insemination (0.620)

on the same data set. They concluded that residual spectrum was not better for diagnosing the pregnancy state. The third strategy of this study was to classify the spectra data to 7 groups based on the period after insemination (1-30 d, 31-60 d, 61-90 d, 91-120 d, 121-150 d, 151-180 d and >180 d). In this part, they only used spectra data after an insemination. The aim of this classification was to reduce pregnancy signal variability as well as spectral variability and therefore enhanced the prediction accuracy. The AUC value of the test set was in a range of 0.590 to 0.650 in class 1 to 5 and increased to 0.700 in class 6 and to 0.820 in class 7. Generally, they concluded that MIR may not be adequate to predict the pregnancy state in early and mid-stages after an insemination. However, the model using records after 150 days of pregnancy showed promising prediction accuracy.

2. Material and Methods

2.1. Data

The data for this study was from the Austrian milk recording for the period of July 2014 to February 2019 and was provided by Zuchtdata GmbH. The test day milk data consisted of information on breed, herd, region, parity, days in milk, milk yield, milk components (fat, protein, urea, lactose), somatic cell count (SCC) and the MIR spectra data for the respective test days. Additionally, we had information about the inseminations and the exact calving dates. Test day records of Fleckvieh, Brown Swiss and Holstein Friesian cows between 3 and 305 days of lactation were included. Information about the pregnancy test IDEXX was also available for some cows. The average lactation day of successful insemination was about day 93. Table 4 shows the number of records of the complete data set.

Table 4: Number of records of the complete data set

Variables	Records
Farms	**6,899**
Animals (Cows)	**40,106**
Fleckvieh	30,589
Brown Swiss	3,854
Holstein Friesian	5,663
Test day records	**403,863**
open	124,163
pregnant	279,700
IDEXX	8,216

The MIR spectra was analysed with infrared spectrometer from the brand Foss, consisting of 1,060 data points. These points are the absorbance values of infrared light at different wavenumbers, with frequencies from 926 to 5010 nm. The received MIR spectra data were standardized into a common basis according to a procedure developed by Grelet et al. (2015). To remove noisy areas, only following spectral areas were selected: 968.1 to 1,577.5 nm, 1,731.8 to 1,762.6 nm, 1,781.9 to 1,808.9 nm and 2,831 to 2,966 nm (Grelet et al. 2016). These selected 212 data points contain most of information whilst the other areas have less information because they are noisy induced by water absorbance or they are not repeatable between the different MIR instruments. According to some other studies (Soyeurt et al. 2011, 2012, Grelet et al. 2016, Lainé et al. 2017, Mineur et al. 2017, Ho et al. 2019, Rienesl et al. 2019) first derivative spectra values (Savitzky-Golay filter) were used for developing the prediction model.

2.2. Data preparation for the final model tests

The different data sets were merged and primary data preparation was done in SAS (SAS Institute Inc., 2017). The pregnancy state of each cow was connected to the associated test day by the following procedure. First, the real gestation length was calculated as the date of re-calving minus the date of the last insemination, which was also defined as the successful date of insemination. If this real gestation length was within an acceptable gestation length range, which is in Fleckvieh between 275 – 305 days, in Holstein between 268 – 298 days and in Braunvieh between 276 – 306 days, the cow and her test days were included, if not, the cow and her test days were deleted. Finally, all test day records before the successful insemination date were coded as 'open' and all test day records between the date of successful insemination and date of re-calving were coded as 'pregnant'. This procedure is visualized for an example in Figure 4. All doubts and irregularities in the recorded dates were removed to ensure the data quality.
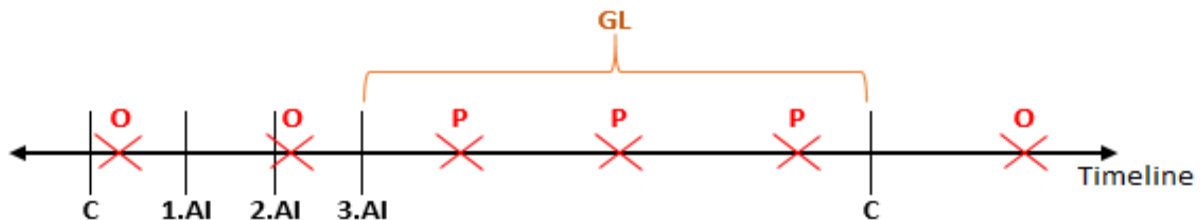
Figure 4: Graphic representation of coding 'open' (O) & 'pregnant' (P) on an example Fleckvieh dairy cow
AI = artificial insemination
C = calving
GL = gestation length (theoretical gestation length in Fleckvieh is between 275 – 305 days)
Red Cross = test day with coding 'open' or 'pregnant'

Another important variable for the prediction model was the pregnancy day of each cow at each test day. This variable was the result of the date of test day minus the date of successful insemination. The next step was the calculation of the first derivative of the MIR spectra data points according to some other studies (Soyeurt et al. 2011, 2012, Grelet et al. 2016, Lainé et al. 2017, Mineur et al. 2017, Ho et al. 2019, Rienesl et al. 2019). The first derivative was calculated by applying the formula $dx(n)=x(n)-x(n+4)$. Finally, the 212 selected wavelengths of the MIR spectra were taken in the final data set (Grelet et al. 2016).

Further data preparations were done in Rstudio (R Development Core Team, 2008). The days in milk (DIM) correction of the 212 selected wavelengths was done according to Vanlierde et al. (2015). For this, each first derivative spectral value was multiplied by a constant (i.e. 1), a linear ($\sqrt{3} * x$) and a quadratic [$\sqrt{5}/4 * (3x^2 - 1)$] modified Legendre polynomial (Gengler et al. 1999), where $x = -1 + 2 [(DIM - 3) / (305 - 3)]$. The range of the application of the equation was not, as in the study of Vanlierde et al. (2015), between 5 and 365 DIM but between 3 and 305 days because of the standard lactation of

305 days in Austria. The results of this modification were 636 (212 data points for each constant, linear and quadratic part) spectral values and they were finally used for the prediction model.

The complete data set consisted of 403,863 records. These records were further randomly split by farm and pregnancy state into half a calibration (training) set and half a validation (test) set (except for the last model test, where also other ratios were split). In this way, it was not possible to have cows from the same herd in calibration and validation set. Additionally, the calibration set got balanced (1:1) in terms of pregnancy states of the cows, thus open or pregnant, by using random down sampling. Further, to evaluate various factors, different settings were applied on calibration sets.

2.3. Calibration and validation settings for final model tests

For the evaluation of various effects in the model, different settings were applied on the calibration subsets of each model test. These different subsets are explained in detail below. In order to get a realistic validation data set, another variable was introduced, the so called 'expected' pregnancy day. This 'expected' pregnancy day was calculated as date of test day minus date of the most recent insemination. This variable was important because in reality, the real pregnancy day is not known and so the 'expected' pregnancy day was used in the validation data set.

2.3.1. Single prediction equation across the whole period

The first step was to develop and apply a single prediction equation across the whole lactation and gestation period, so there was one prediction equation for every test day record, regardless of the lactation and gestation stage. Further, the data set with all cows was compared to a data set with cows with only one insemination. The total number of records for each data set is given in Table 5.

Application of a single prediction equation on all cows

First we applied a single prediction equation with no restrictions in the calibration and also the validation data set to all cows in the data set. In this step, all 403,863 records (Table 5) were used and were randomly split into half a calibration and half a validation data set. To have the same number of pregnant and open cows in the calibration data set, the random down sampling was used again. The validation data set was kept unbalanced to get realistic conditions.

Application of a single prediction equation on cows with only one insemination

In this part, the prediction equation was used on animals with only one insemination. This step was done because of the relatively high values of pregnancy losses at the beginning of the gestation, which is mentioned in section 1.3.2. For animals with only one insemination and confirmed next calving, non-pregnancy and pregnancy may be determined without error. In total there were 225,229 test day records of cows, which were inseminated only once before calving. These records were randomly split 1:1 into calibration and validation set. Additionally, to get the same numbers of pregnant and open cows in the calibration data set, random down sampling was used. The exact ratio of pregnant and open animals can be seen in Table 5. The average lactation day of the successful insemination in this class was about day 70. The validation set was kept unbalanced.

Table 5: Number of records of cows with one insemination at all and number of all records

	Records of animals with one insemination	Records of all animals
open	50,775	124,163
pregnant	174,454	279,700
total	225,229	403,863

2.3.2. A prediction equation for each different 'expected' pregnancy stage

After the first results, we decided to make different prediction equations for different 'expected' pregnancy stages. The different stages, which were used, are given in Table 6. In the calibration set, the real pregnancy day of a cow at any test day was used for the classification. So, in the training set there were only animals which were certainly in the respective gestation stage. However, in the test set the 'expected' pregnancy day was used to classify the animals to the respective stage. In terms of the open animals, there were no differences between the calibration and validation data set. In each class, all open animals were used, which were not inseminated yet and also the open animals with a non-successful insemination and an 'expected' pregnancy day in the respective pregnancy class.

*Table 6: Subclasses depending on the pregnancy day in calibration set and
'expected' pregnancy day in validation set*

Pregnancy class	Calibration setting *pregnancy day*	Validation setting *'expected' pregnancy day*
1	1 - 30	1 - 30
2	31 - 60	31 - 60
3	61 - 90	61 - 90
4	91 - 120	91 - 120
5	121 - 150	121 - 150
6	151 - 305	151 - 305

2.3.3. A prediction equation for each different 'expected' pregnancy stage and lactation stage

The next objective was to evaluate the effects of the classification into different 'expected' pregnancy stages and lactation stages to the accuracy of the prediction model. For this step, different prediction equations for different 'expected' pregnancy and lactation stages were produced. The individual classes are presented in Table 7. In the calibration set, the real pregnancy day and the lactation day were used to classify the test day record to the associated prediction equation. Additionally, in some training settings only cows with one insemination were used. It was not possible to use only animals with one insemination in all calibration settings because of the small numbers of records in some classes. The respective setting for each class is also given in Table 7. The code 'O1P1' means that there were open and pregnant animals with one insemination. 'OAP1' means that there were pregnant animals with one insemination and all open cows. The last code 'OAPA' means, that there were no restrictions concerning to the number of inseminations of open and pregnant cows. In the validation data set, the 'expected' pregnancy day and the lactation day were used for the classification. No further settings or restrictions were applied to get a realistic validation data set. In this part also the open cows were classified into these classes in both data sets.

Table 7: Subclasses depending on the pregnancy day & lactation day in calibration set and 'expected' pregnancy day & lactation day in validation set

	Calibration setting		Validation setting
	pregnancy day	setting code	*'expected' pregnancy day*
1 - 60	1 - 30	O1P1	1 - 30
	31 - 60	O1P1	31 - 60
61 - 120	1 - 30	O1P1	1 - 30
	31 - 60	O1P1	31 - 60
	61 - 90	O1P1	61 - 90
	91 - 120	O1P1	91 - 120
121 - 180	1 - 30	OAP1	1 - 30
	31 - 60	OAP1	31 - 60
	61 - 90	OAP1	61 - 90
	91 - 120	OAP1	91 - 120
	121 - 150	OAP1	121 - 150
	151 - 180	OAP1	151 - 180
181 - 240	1 - 30	OAPA	1 - 30
	31 - 60	OAPA	31 - 60
	61 - 90	OAP1	61 - 90
	91 - 120	OAP1	91 - 120
	121 - 150	OAP1	121 - 150
	151 - 180	OAP1	151 - 180
	181 - 210	OAP1	181 - 210
	211 - 240	OAP1	211 - 240
241 - 305	1 - 30	OAPA	1 - 30
	31 - 60	OAPA	31 - 60
	61 - 90	OAPA	61 - 90
	91 - 120	OAPA	91 - 120
	121 - 150	OAP1	121 - 150
	151 - 180	OAP1	151 - 180
	181 - 210	OAP1	181 - 210
	211 - 240	OAP1	211 - 240
	241 - 305	OAP1	241 - 305

O1P1 = open and pregnant animals with one insemination
OAP1 = all open animals and pregnant animals with one insemination
OAPA = all open and pregnant animals

2.3.4. Effect of different sample sizes for the calibration set

Another objective of this study was to evaluate the effects of different sample sizes of calibration data sets to the accuracy of the prediction model. As already mentioned above, the records were normally split randomly half in calibration and half in validation data set. To evaluate the effects, the sample sizes of the calibration data set for the different sub settings were reduced as presented in Table 8.

Table 8: Sample size classes of the calibration data set

Class	Calibration		Class	Calibration	
	Open	Pregnant		Open	Pregnant
10	10	10	1,000	1,000	1,000
20	20	20	2,000	2,000	2,000
50	50	50	5,000	5,000	5,000
100	100	100	10,000	10,000	10,000
200	200	200	20,000	20,000	20,000
500	500	500	all	all	all

The validation data set consisted in each sub setting of half of the full data set.

2.4. Methodology

Model predictions were done with Partial Least Square Analysis (PLS) by applying the function 'trainControl' of R package 'caret' (Kuhn 2008). A 10-fold cross validation was used, the number of components was set automatically for every run and discrimination was done by class probabilities. The spectra values were centered and scaled for prediction.

According to the sample size generator (https://www.stat.ubc.ca/~rollin/stats/ssize/n2.html), 5 replications per setting were used for the final model tests. A standard deviation of 0.003 was given for replicates and it allowed to detect significance at p-value of 0.05 for differences of around 0.007.

The indicators of model fit were as follows (Lantz 2015):

1. Sensitivity: is defined as the proportion of pregnant cows predicted correctly as pregnant
2. Specificity: is defined as the proportion of open cows predicted correctly as open
3. Balanced Accuracy: is defined as the mean of sensitivity and specificity
4. AUC values (Area under the ROC curve):

The ROC (Receiver Operating Characteristic) curve is defined as the plot of the true positive rate TPR (=sensitivity) versus the false positive rate FPR (= 1 – specificity) for different values of the threshold, so it is a graph showing the performance of a classification. An example is displayed in Figure 5. The diagonal line represents random classification. A ROC curve above the diagonal line means that a prediction model is better than the random classification and a ROC curve under the diagonal line means that the prediction is worse. The AUC value indicates the accuracy of a model and it can be classified into classes, which can be seen in Table 9.

Table 9: Classification of AUC values (according to Lantz 2015)

AUC value	Classification
> 0.9	outstanding
0.8 to 0.9	excellent/good
0.7 to 0.8	acceptable/fair
0.6 to 0.7	poor
0.5 to 0.6	no discrimination
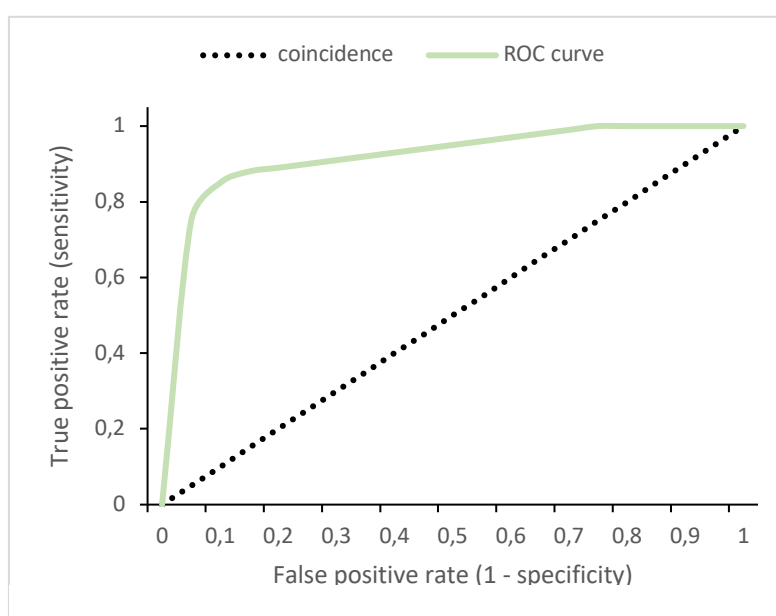
Figure 5: Example for a ROC curve (according to Lantz 2015)

3. Results

3.1. Application of a single prediction equation across the whole period

Results for the application of a single prediction equation across the whole period for the data set of cows with only one insemination and the data set with all cows are shown in Table 10. This Table includes the indicators of model fit (sensitivity, specificity and balanced accuracy) for the large, imbalanced validation set for each model.

Table 10: Mean and standard deviation (sd) of sensitivity, specificity and balanced accuracy of the model with animals with only one insemination and the model with all animals

	sensitivity		specificity		balanced accuracy	
	mean	sd	mean	sd	mean	sd
one insemination	0.889	0.002	0.904	0.003	0.896	0.001
all inseminations	0.856	0.003	0.836	0.004	0.846	0.001

In general, all three indicators of model fit were higher in the model which used only cows with one insemination. The sensitivity increased from 0.856 in the formula on all cows to 0.889 in the formula on cows with one insemination. The specificity increased from 0.836 to 0.904. For this reason, the final calibration data sets consisted, when it was possible, of animals which were only inseminated once. The test (validation) data set still contained all animals.

The sensitivity, using the formula across the whole period on all animals, for individual pregnancy months are displayed in Table 11. The sensitivity in the first pregnancy month was about 0.380, increased in the second pregnancy month to about 0.695, in the third pregnancy month to about 0.945 and from the fourth month the sensitivity was above 0.999.

Table 11: Sensitivity of the single prediction equation on all animals divided into the different pregnancy months

pregnancy month	sensitivity	pregnancy month	sensitivity
1	0.380	6	0.999
2	0.695	7	1.000
3	0.945	8	0.999
4	0.997	9	1.000
5	0.999	10	1.000

3.2. A prediction equation for each different 'expected' pregnancy stage

Because of the results in chapter 3.1, the next step was to create different prediction equations for each different expected pregnancy stage, as explained in chapter 2.3.2. The results of this classification into pregnancy stages is given in Table 12.

Table 12: Mean and standard deviation of sensitivity, specificity, balanced accuracy and AUC value of the model with different 'expected' pregnancy classes

'expected' pregnancy stage	sensitivity		specificity		balanced accuracy		AUC	
	mean	sd	mean	sd	mean	sd	mean	sd
1 - 30	0.825	0.003	0.679	0.003	0.752	0.001	0.818	0.001
31 - 60	0.912	0.007	0.838	0.003	0.875	0.003	0.929	0.001
61 - 90	0.964	0.003	0.923	0.003	0.943	0.001	0.973	0.001
91 - 120	0.986	0.001	0.959	0.001	0.972	0.001	0.988	0.001
121 - 150	0.995	0.001	0.978	0.001	0.986	0.001	0.995	0.001
151 - 305	0.996	0.001	0.991	0.001	0.993	0.001	0.999	0.001

This model test showed that the indicators for model fit increased with a higher pregnancy stage. Comparing those results with the results applying one prediction equation across the whole lactation (see Table 11), the sensitivity for the first pregnancy month increased from about 0.380 in the formula across the whole period to about 0.825 in the formula for the first pregnancy stage. Regarding specificity, the mean value decreased from 0.836 through the whole period to 0.679 in the formula for the first pregnancy stage. In the second month, the sensitivity increased again from 0.695 in the formula across the whole period to 0.912 in the formal for the second pregnancy month and the specificity increased also from about 0.825 in the formula across the whole period to 0.838 in the formula for the second pregnancy stage. In the third pregnancy month, the indicators were also higher in the formula with the different 'expected' pregnancy stages. In the fourth month, the sensitivity was higher in the formula across the whole period (0.997) than in the formula for the fourth pregnancy stage (0.986) but the specificity was higher, 0.959 compared with 0.836. The results of the formula for the last two 'expected' pregnancy stages were nearly the same like in the formula across the whole period concerning to sensitivity. Specificity was higher in the formula for the different pregnancy stages.

Table 13 displays the results of the formulas for different 'expected' pregnancy stages depending on the lactation stages. It is shown that the sensitivity increased with the lactation month and the specificity decreased with the lactation month. There were irregularities only in the last lactation

month. For this reason, formulas for each different 'expected' pregnancy stage and also lactation stage were created.

Table 13: Sensitivity and specificity of the different 'expected' pregnancy stages depending on lactation stages

lact.	'expected' pregnancy stage											
mo.	1 - 30		31 - 60		61 - 90		91 - 120		121 - 150		151 - 305	
	sens.	spec.	sens.	spec.	sens.	spec.	sens.	spec.	sens.	spec.	sens.	spec.
1	0.000	0.999	-	-	-	-	-	-	-	-	-	-
2	0.124	0.965	0.000	0.999	-	-	-	-	-	-	-	-
3	0.840	0.276	0.513	0.855	0.043	0.997	-	-	-	-	-	-
4	0.995	0.011	0.947	0.158	0.766	0.641	0.345	0.960	-	-	-	-
5	0.997	0.006	0.996	0.016	0.983	0.077	0.916	0.370	0.784	0.826	-	-
6	0.996	0.006	0.995	0.006	0.998	0.006	0.996	0.028	0.966	0.228	0.689	0.840
7	0.991	0.018	1.000	0.004	1.000	0.000	0.999	0.000	0.999	0.000	0.952	0.341
8	0.955	0.086	0.997	0.004	1.000	0.000	1.000	0.000	0.999	0.000	0.999	0.009
9	0.784	0.330	0.984	0.019	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
10	0.589	0.496	0.743	0.216	0.973	0.067	0.994	0.000	0.999	0.000	0.999	0.000

lact. mo. = lactation month
sens. = sensitivity
spec. = specificity

3.3. A prediction equation for each different 'expected' pregnancy and lactation stage

Based on the results in chapter 3.2, the final step was to create prediction equations for each different 'expected' pregnancy and also lactation stage, as explained in chapter 2.3.3. The results of this classification are provided in Table 14. The sensitivities were in a range of 0.494 and 0.995, thus, differences were very big. The range of specificities was also very big, from 0.512 to 0.884.

Within lactation stage, the indicators of model fit increased with 'expected' gestation stage. So for example, in the first lactation stage, which was from day 1 to day 60 of lactation, the sensitivity increased from 0.890 for the first 'expected' gestation stage, which was between 1 and 30 days of expected pregnancy day, to 0.946 in the second 'expected' gestation stage, which was between 31 and 60 days of expected pregnancy day. However, the standard deviation was higher in the second gestation stage because of the low number of 'pregnant' records in this stage.

Regarding to the 'expected' gestation stages, the indicators for model fit were mostly decreasing with lactation stages. Sensitivities for the first 'expected' gestation stage (1 to 30 days of 'expected' pregnancy) were 0.890 in the first lactation stage, 0.620 in the second lactation stage, and 0.494 in the third lactation stage.

Table 14: Mean and standard deviation of sensitivity, specificity, balanced accuracy and AUC value of the model with different 'expected' pregnancy and lactation classes

lactation stage	records test set		sensitivity		specificity		balanced accuracy		AUC		'expected' pregnancy stage
	open	pregnant	mean	sd	mean	sd	mean	sd	mean	sd	
1 - 60	35,158	2,497	0.890	0.007	0.690	0.009	0.790	0.003	0.853	0.002	1 - 30
1 - 60	32,479	37	0.946	0.090	0.750	0.024	0.848	0.040	0.919	0.046	31 - 60
61 -120	15,023	11,421	0.620	0.007	0.512	0.003	0.566	0.003	0.592	0.003	1 - 30
61 -120	10,992	9,575	0.827	0.009	0.623	0.009	0.725	0.001	0.792	0.001	31 - 60
61 -120	8,780	2,438	0.922	0.005	0.840	0.004	0.881	0.003	0.937	0.002	61 - 90
61 -120	8,388	43	0.969	0.045	0.857	0.056	0.913	0.014	0.974	0.013	91 - 120
121 - 180	3,647	4,303	0.494	0.011	0.553	0.014	0.523	0.003	0.534	0.005	1 - 30
121 - 180	2,819	7,791	0.565	0.008	0.590	0.017	0.577	0.008	0.606	0.010	31 - 60
121 - 180	2,037	11,555	0.597	0.008	0.625	0.002	0.611	0.003	0.658	0.004	61 - 90
121 - 180	1,318	9,074	0.772	0.005	0.676	0.008	0.724	0.005	0.798	0.004	91 - 120
121 - 180	1,324	2,367	0.897	0.006	0.834	0.007	0.866	0.004	0.929	0.005	121 - 150
121 - 180	1,157	45	0.951	0.050	0.884	0.032	0.918	0.014	0.964	0.013	151 - 180
181 - 240	990	1,582	0.553	0.014	0.527	0.015	0.54	0.003	0.559	0.005	1 - 30
181 - 240	771	2,729	0.592	0.014	0.597	0.023	0.594	0.007	0.636	0.011	31 - 60
181 - 240	573	4,606	0.607	0.015	0.622	0.024	0.614	0.006	0.66	0.007	61 - 90
181 - 240	459	7,918	0.587	0.014	0.652	0.028	0.620	0.009	0.666	0.009	91 - 120
181 - 240	380	11,407	0.623	0.013	0.623	0.032	0.623	0.014	0.674	0.010	121 - 150
181 - 240	288	9,506	0.766	0.010	0.718	0.017	0.742	0.011	0.821	0.013	151 - 180
181 - 240	289	2,477	0.896	0.010	0.846	0.015	0.871	0.005	0.939	0.005	181 - 210
181 - 240	279	48	0.995	0.011	0.820	0.006	0.908	0.006	0.958	0.015	211 - 240
241 - 305	307	559	0.504	0.024	0.584	0.036	0.545	0.016	0.568	0.015	1 - 30
241 - 305	256	938	0.598	0.018	0.537	0.039	0.568	0.012	0.612	0.016	31 - 60
241 - 305	155	1,643	0.623	0.012	0.633	0.038	0.628	0.020	0.675	0.026	61 - 90
241 - 305	155	2,759	0.610	0.011	0.706	0.023	0.658	0.009	0.713	0.011	91 - 120
241 - 305	107	4,557	0.612	0.009	0.676	0.046	0.644	0.021	0.691	0.015	121 - 150
241 - 305	93	8,153	0.638	0.013	0.725	0.042	0.681	0.022	0.742	0.021	151 - 180
241 - 305	88	11,741	0.684	0.013	0.768	0.056	0.726	0.026	0.792	0.018	181 - 210
241 - 305	83	8,563	0.776	0.007	0.849	0.049	0.813	0.022	0.885	0.019	211 - 240
241 - 305	79	506	0.852	0.015	0.855	0.036	0.854	0.021	0.933	0.013	241 - 305

3.4. Effect of different sample sizes for the calibration set

For this model test, the sample sizes for the different calibration subsets were reduced to the numbers, which are displayed in Table 8 in chapter 2.3.4. The results are shown in Table 15.

Table 15: Effect of different sample sizes for the calibration set

sample size train	all animals					
	sensitivity		specificity		bal. acc.	
	mean	sd	mean	sd	mean	sd
10	0.801	0.107	0.838	0.066	0.822	0.023
20	0.849	0.078	0.811	0.063	0.83	0.012
50	0.847	0.014	0.825	0.014	0.836	0.004
100	0.856	0.019	0.817	0.025	0.837	0.007
200	0.862	0.026	0.818	0.027	0.84	0.001
500	0.848	0.015	0.834	0.013	0.841	0.001
1000	0.853	0.012	0.831	0.013	0.842	0.001
2000	0.857	0.01	0.829	0.01	0.843	0.002
5000	0.857	0.003	0.831	0.003	0.844	0.001
10,000	0.857	0.002	0.834	0.003	0.845	0.001
20,000	0.853	0.002	0.839	0.002	0.846	0.001
all	0.856	0.003	0.836	0.004	0.846	0.001

The differences between the different sample size classes were not that big but the standard deviation was higher in the small sample size classes.

4. Discussion

4.1. Comparison of the single prediction equation across the whole period applied on data of cows with one insemination and on all cows

Due to the results of the first part of this work (section 3.1), the results of the single prediction equation on cows with only one insemination were in all indicators for model fit (sensitivity, specificity, balanced accuracy) significantly higher, compared to the results of the single prediction equation with all cows. The results are visualized in Figure 6.

However, in practice, there are not only cows with one insemination, so the general use of a single prediction equation on cows with one insemination is not realistic. But the use of cows with only one insemination in the calibration data set could increase the indicators for model fit. So for this reason, in the final model, which was subdivided into the different pregnancy and lactation stages, in calibration we only used cows with one insemination, when the number of records was high enough in this group.

The comparison of these two data sets also showed that in the prediction equation on cows with only one insemination the specificity was higher than the sensitivity and in the formula on all cows it was vice versa. With regard to the standard deviation, there were no significant differences between the different groups.
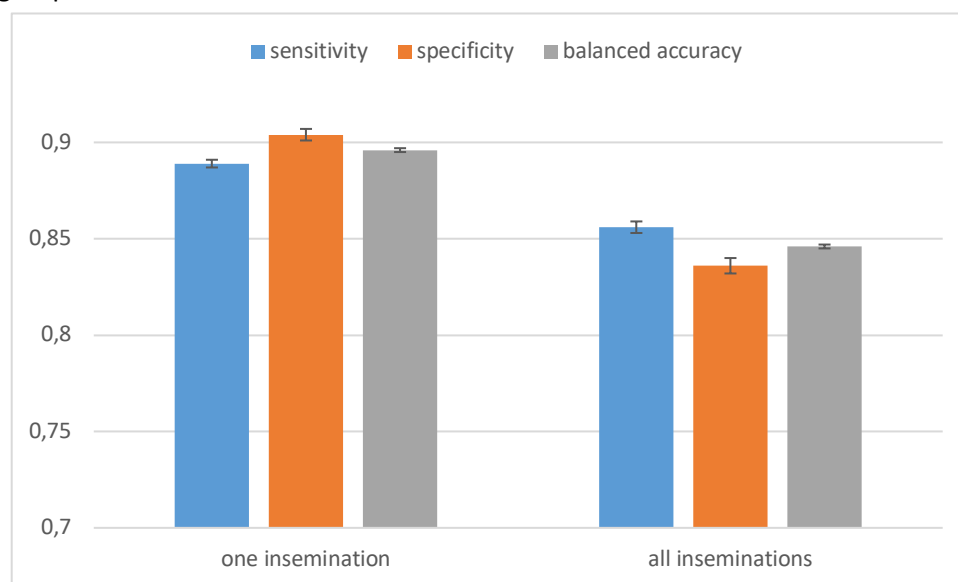

Figure 6: Sensitivity, Specificity and balanced accuracy of the formula on cows with only one insemination and the formula on cows with all inseminations

According to the study of Laine et al. (2013), which used also all cows with no restrictions compared to the number of inseminations, the sensitivity (>0.99) was much higher than in this work (0.856). One reason for these higher sensitivities could be the observed time range. In the study of Laine et al. (2013)

only the first 50 days after an insemination were observed and, in this study, respectively in this part of the study, the whole period was observed. With regard to the specificity, the differences were not that big: 0.862 (Laine et al. 2013) compared to 0.836 in this study. Another reason could also be that Laine et al. (2013) only used a residual spectrum and in this work the raw milk MIR spectrum was used.

As already mentioned, in the study of Delhez et al. (2020) there are two different strategies related to one formula across the whole period. In the first strategy, which used only spectral records after an insemination, the sensitivity (0.650) and also specificity (0.560) were much lower than in this work (0.856, 0.836). Also in the second part, which considered the spectral differences, the sensitivity (0.590) and specificity (0.520) were much lower. One reason for these differences could be the different strategies of the use of MIR spectra to predict the pregnancy state. Another reason could be the lower number of records. The study of Delhez et al. (2020) included information about MIR spectra and inseminations from 8,064 Holstein cows and this study included information from 40,106 Fleckvieh, Holstein and Brown Swiss cows.

According to the results in Table 11, which showed the sensitivity of the single prediction equation on all cows and additionally separated by the expected pregnancy month, the sensitivity was very low in the first pregnancy month and increased with pregnancy months. The course of these sensitivity values and the percentage of test day records in each different pregnancy month is displayed in Figure 7. The relatively high value of sensitivity across the whole period (0.856) was possible because only about 29 % of the total records were in the first and second pregnancy month. So, in the first two pregnancy months, the prediction was not that good but in the other months the prediction was >0.94. For this reason of low values during the beginning of gestation, the formula was split into different pregnancy stages.
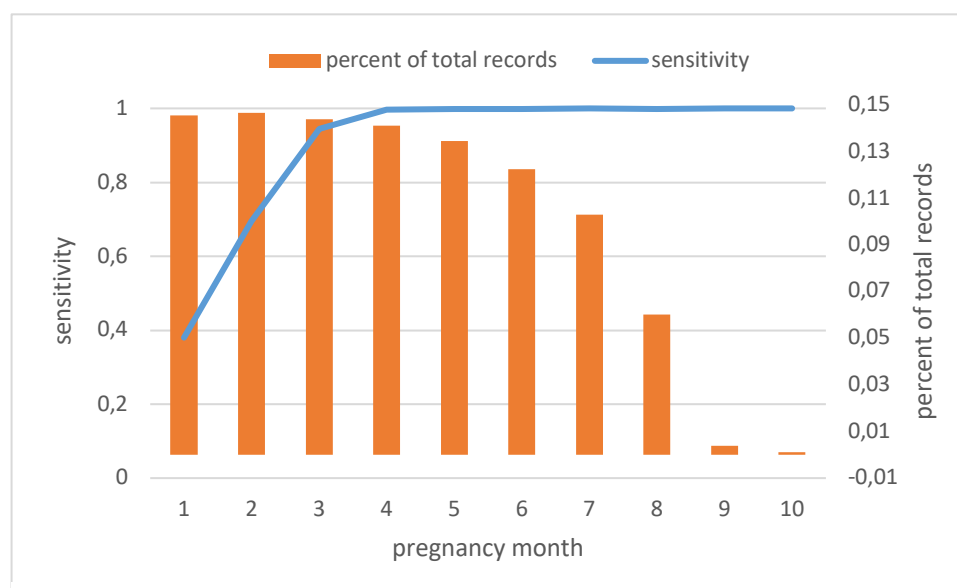

Figure 7: Course of the sensitivity value of the single prediction equation on all cows additionally separated into the different expected pregnancy months and the distribution of the test day records

4.2. A prediction equation for each different 'expected' pregnancy stage

In general, the indicators for model fit increased with a higher pregnancy stage. Comparison of the single prediction equation across the whole period and prediction equations for each different 'expected' pregnancy stage showed some interesting differences. First, especially the sensitivity in the first two months of 'expected' pregnancy was different. This can also be seen in Figure 8. The sensitivity increased from 0.380 in the single prediction equation across the whole period to about 0.825 in the prediction equation for the first 'expected' pregnancy stage. Also, in the second pregnancy month, the sensitivity increased significantly from 0.695 to 0.912. From the third pregnancy month on, the sensitivity was nearly the same in both formulas.

Regarding to the specificity, there was a lower value for the first pregnancy month in the prediction equation for the first 'expected' pregnancy month (0.679) compared to the single prediction equation across the whole period (0.836). In the second pregnancy month, the specificity was approximately the same in the two models and from the third pregnancy month on, also the specificity was higher in the model with 'expected' pregnancy classes.
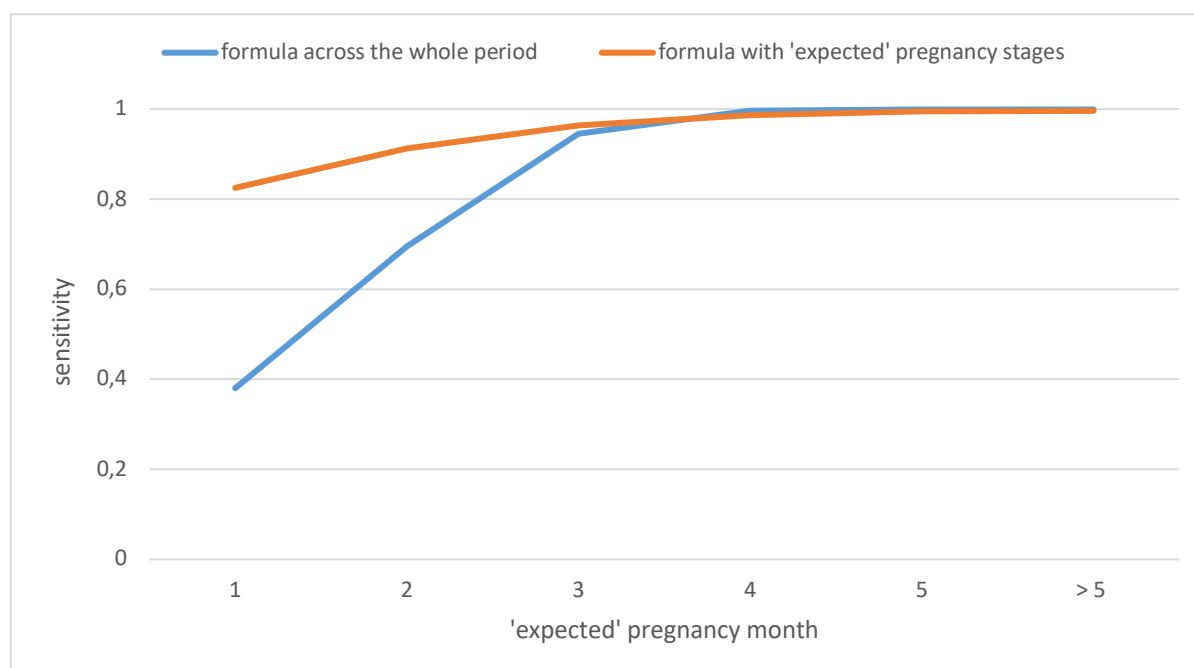

Figure 8: Sensitivity of the model across the whole period, additionally separated by pregnancy month, and the model with pregnancy stages

Sensitivity and specificity were more balanced in the model with different prediction equations for each different 'expected' pregnancy stage than in the model across the whole period, especially from the third pregnancy month on. The balanced accuracies were always higher in the models with different 'expected' pregnancy classes. This is visualized in Figure 9.

According to the study of Delhez et al. (2020), which also made different classes after insemination, the indicators of model fit are much better in this study. Sensitivity and specificity of the first 30 days after insemination were at 0.570 and 0.580 in the study of Delhez et al. (2020). In this study, the sensitivity was 0.825 and the specificity was 0.679. The same differences can be observed in the other classes. One reason for these differences could be the fact that Delhez et al. (2020) only used records after an insemination and in this study records before and after an insemination were used. Another reason could be the different number of records per class.
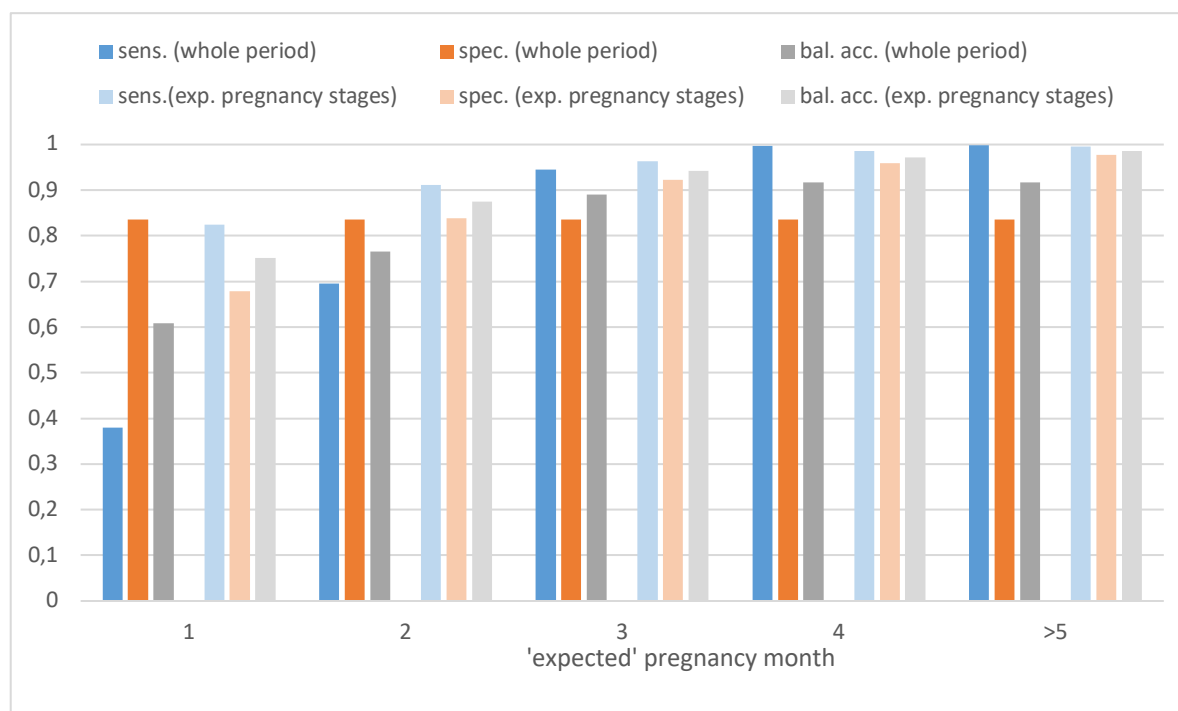

Figure 9: Sensitivity, specificity and balanced accuracy of the formula across the whole period and for the formula with pregnancy stages
exp. = expected
sens. = sensitivity
spec. = specificity
bal. acc. = balanced accuracy

Additionally, analyses and prediction equations were separated by month of lactation. The results of this separation are displayed in Table 13. Results indicate that during the first months of lactation, the sensitivity was very low and the specificity was very high. At the end of lactation, it was visa versa. During the different lactation months, in the formula for the first 'expected' pregnancy stage, the values of sensitivities increased to over 0.99 in lactation month 4 and the specificity decreased to approximately 0.011 in the same month of lactation. This example can be seen in Figure 10.

The same results can also be observed with regard to the other pregnancy months. At the beginning of the lactation, the majority of the animals were predicted as 'open' and with increasing of the lactation day, more animals were classified as 'pregnant'. Overall, the indicators for model fit were

getting better but when these results were separated by month of lactation, the results were similar to the results of the example for the first pregnancy month, which are explained above.
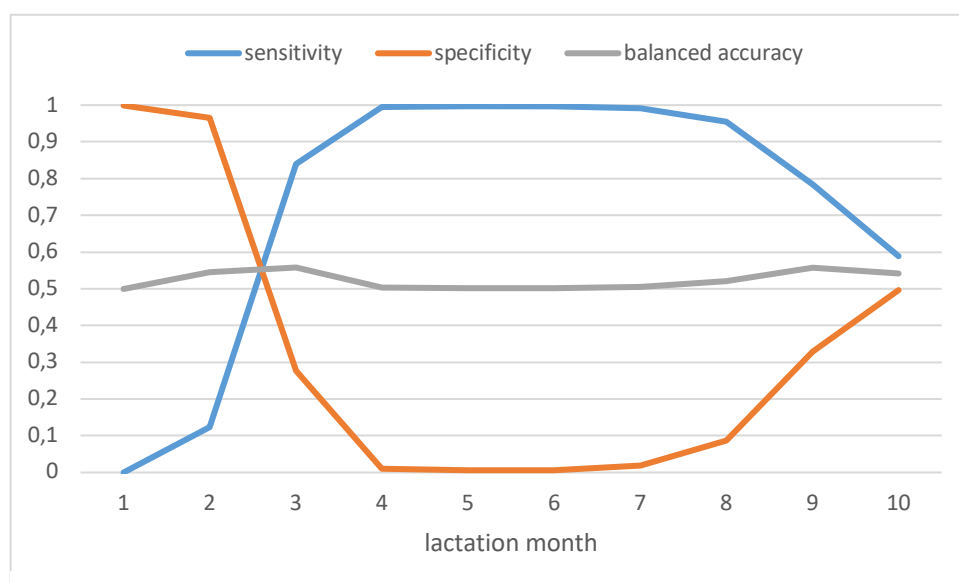

Figure 10: Sensitivity, specificity and balanced accuracy of the formula for the first 'expected' pregnancy month separated into the different lactation months

4.3. A prediction equation for each different 'expected' pregnancy stage and lactation stage

According to the results of the first two parts of this work (section 3.1, 3.2, 4.1, 4.2), we created prediction equations for each different 'expected' pregnancy stage and also lactation stage in the final model test. The results of this final model test are given in Table 14. The comparison of the model with different prediction equations for each different 'expected' pregnancy stages and the model with different prediction equations for each different 'expected' different pregnancy stages and lactation stages showed that the balanced accuracy was getting better in the final model with 'expected' pregnancy and lactation stages. Figure 11 displays the results of the final model concerning to the indicators of model fit for the first pregnancy month. The sensitivity in the formula for only pregnancy stages was in average 0.062 for the first two months of lactation and increased in the final model to 0.890. On the other hand, the specificity in the prediction equation for only pregnancy stages was in average 0.982 for the first two months of lactation and decreased in the final model to 0.690. The sensitivity was higher in the final model but the specificity was lower than in the model with only 'expected' pregnancy month. However, the balanced accuracy was also higher in the final model and increased from 0.522 to 0.790.
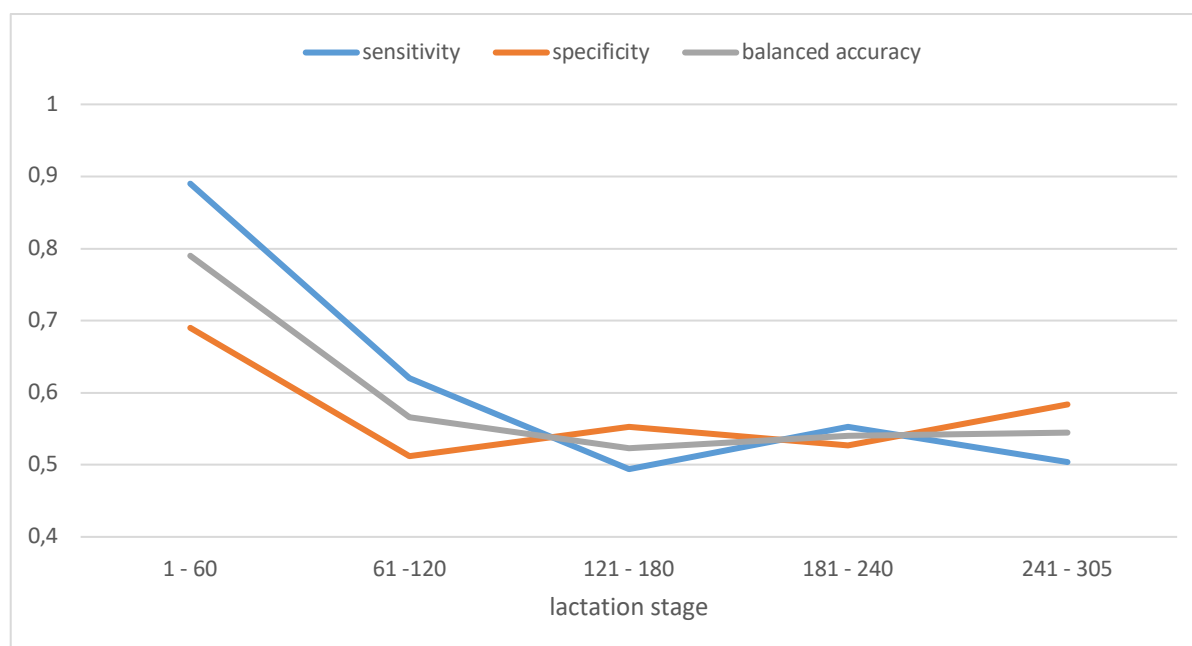
Figure 11: Sensitivity, specificity and balanced accuracy of the formula of the final model for the first 'expected' pregnancy month and in the different lactation stages

According to the results of the final model (Table 14), the following two trends was observed. Within a lactation stage, the values of the indicators of model fit are increasing with 'expected' pregnancy stage. This trend was observed in each lactation stage. These promising prediction accuracy with the increasing day after insemination was also shown in the study of Delhaz et al. (2020). The second trend was that the indicators of model fit decreased with the lactation day of successful insemination.

Three examples for the application of the final model are described here:

As already mentioned, the day of successful insemination was on average lactation day 93. So at first we assumed, that a cow got pregnant at this time and there was also a test day record at this point. A second example was made with an animal, which got pregnant 50 % earlier than the average (lactation day 47), and a third example with an animal which got pregnant 50 % later than the average (lactation day 140). Furthermore, we assumed that there was a test day every 35 days. The exact example setting can be seen in Table 16. Finally, each test day was assigned to the associated formula of the final model. The results of the example are displayed in Figure 12. The best results can be observed in the example with the earliest date of successful insemination (example 2). In this example the balanced accuracy was 0.79 at the beginning of the pregnancy regard to 0.566 in the example of an average date of successful insemination and 0.523 in the example of a latest date of successful insemination.

Table 16: The setting of three different examples for the application of the final model

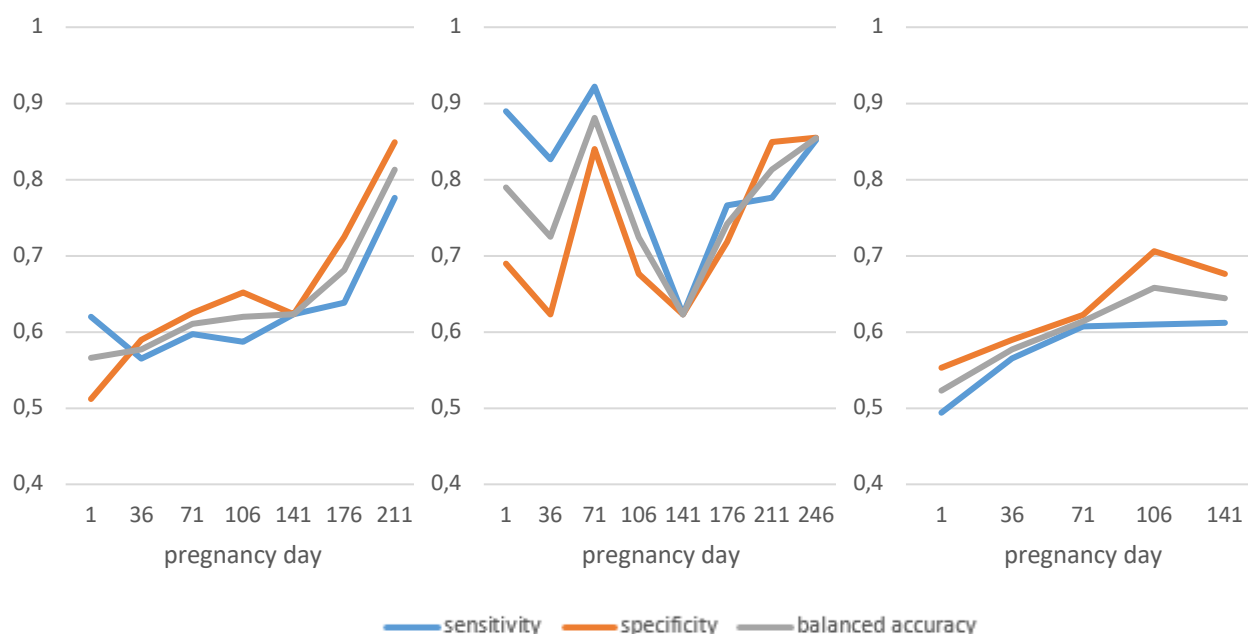| Testday	Day of lactation			Pregnancy day		
	Example 1	Example 2	Example 3	Example 1	Example 2	Example 3
1	93	47	140	1	1	1
2	128	82	175	36	36	36
3	163	117	210	71	71	71
4	198	152	245	106	106	106
5	233	187	280	141	141	141
6	268	222	-	176	176	-
7	303	257	-	211	211	-
8	-	292	-	-	246	-

Figure 12: Sensitivity, specificity and balanced accuracy of the final model applied on three examples
Example 1: date of successful insemination was lactation day 93 (average of the data set)
Example 2: date of successful insemination was lactation day 47 (- 50 % of the average of the data set)
Example 3: date of successful insemination was lactation day 140 (+ 50 % of the average of the data set)

The application of the final model on three real cows is given in Figures 13, 14 and 15. The timeline with the date of the different events, like calving (C) or insemination (AI), is on the x-axis and on the y-axis are the probabilities of pregnancy, which the final model predicated.

The cow in Figure 13 was inseminated twice and the date of successful insemination was on lactation day 52. The pregnancy prediction at the first test day was wrong because the cow was not pregnant at this time. The other test days were after the successful insemination and so the cow was pregnant at these test days. But there were also two test days (4. & 6.) which were predicted incorrectly. The highest probability of pregnancy was at test day 3 with a value of about 77 %.

The second real example is shown in Figure 14 and the cow was inseminated three times and the successful insemination was at lactation day 125. It can be seen that all test days were predicted as 'open' but this was only correct for the first and second test days.

The third example in Figure 15 depicts a cow inseminated once and the day of successful insemination was lactation day 70. In this example, there was also a positive IDEXX test result available. The prediction was correct on all test days and the probabilities were in a range of 58 % to 76 %.
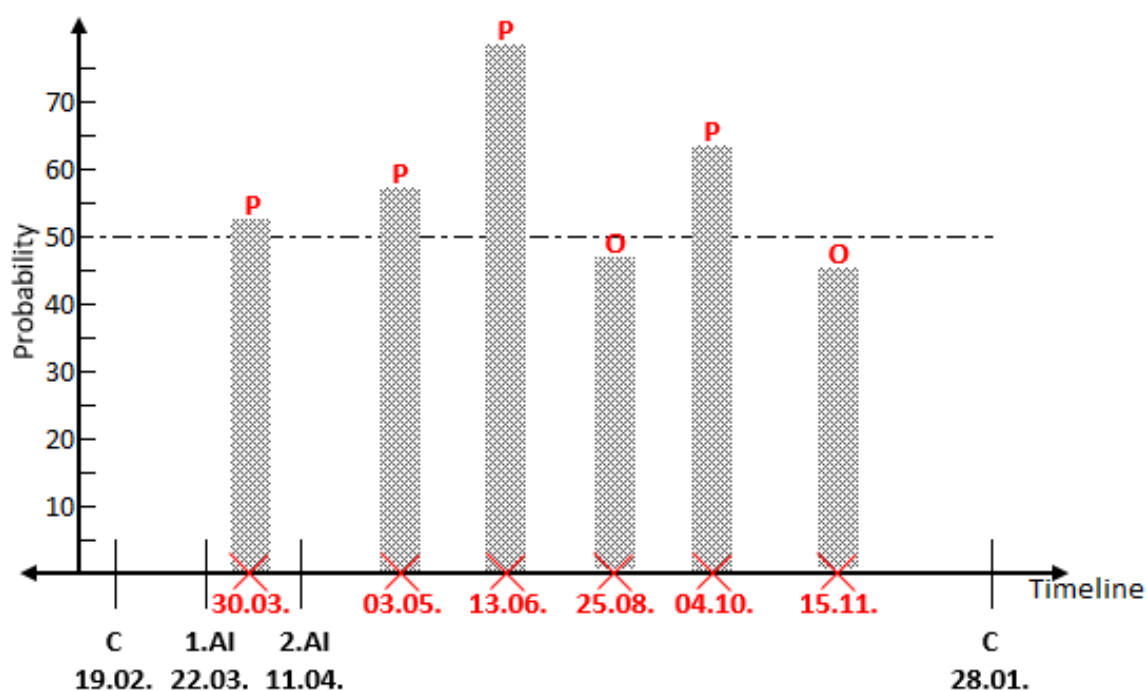

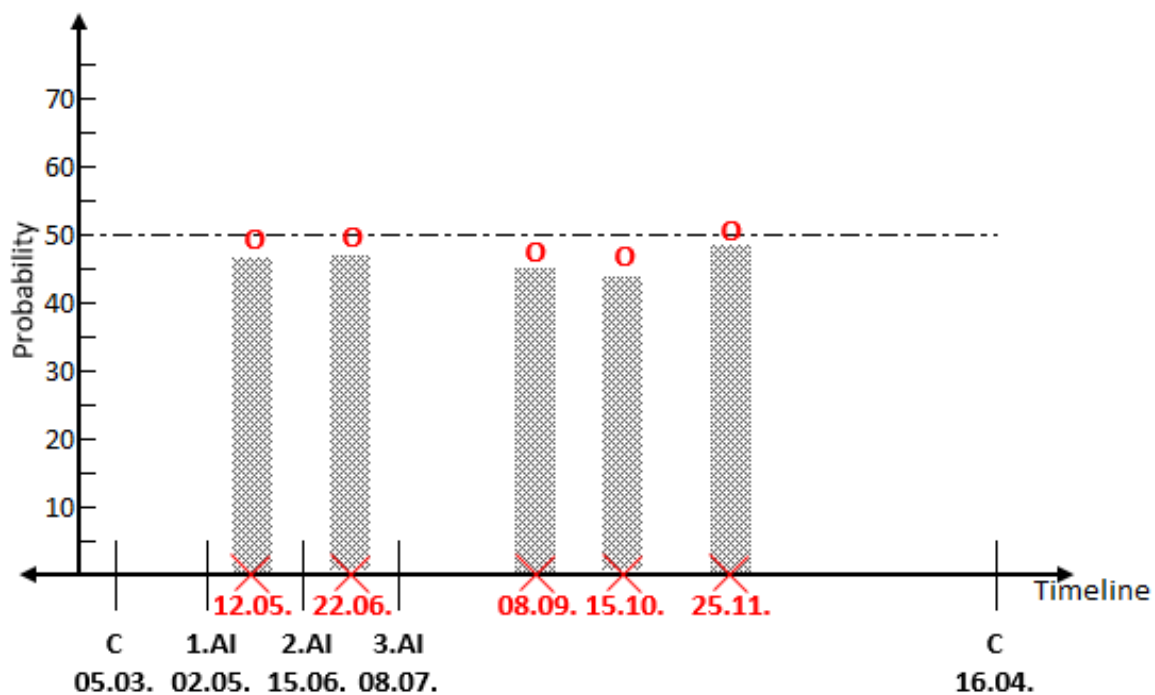*Figure 13: Application of the final model to a real cow (example 1)*

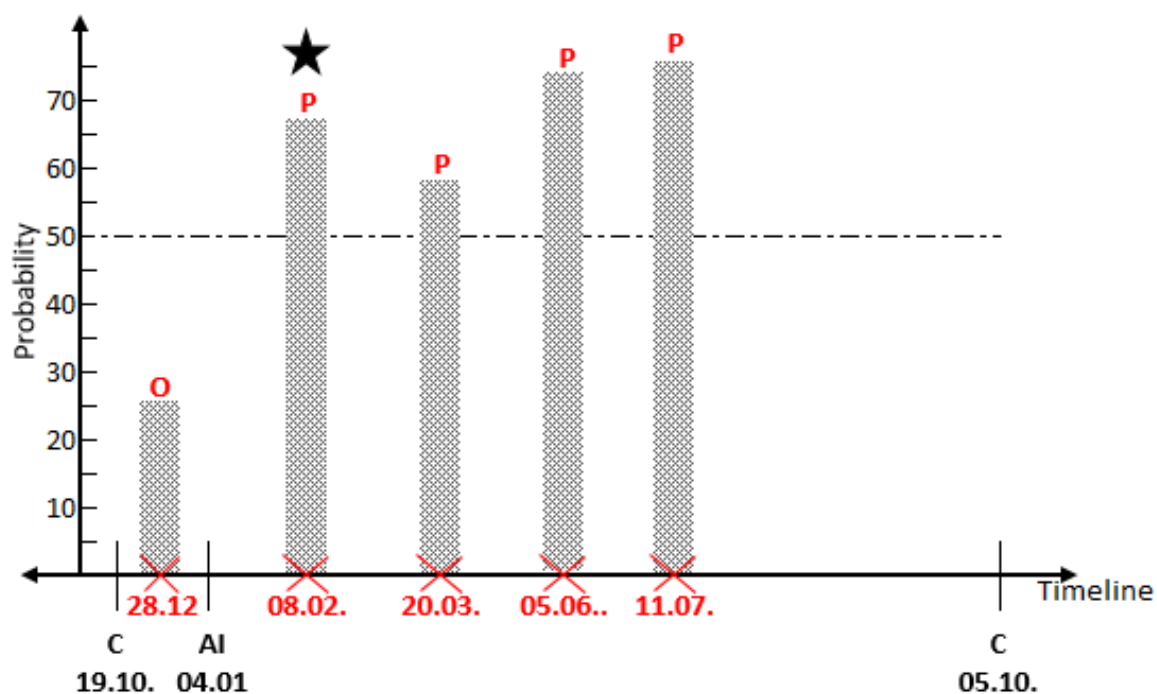Figure 14: Application of the final model to a real cow (example 2)

Figure 15: Application of the final model to a real cow (example 3)

4.4. Effect of different sample sizes for calibration set

The effect of different sample sizes of the calibration set was the last objective of this thesis. For this part, the results will be discussed and compared with the study of Saccenti and Timmerman (2016) and the study of Rienesl et al. (2019). These papers are two of the very rare references on sample size determination.

Just like in this study, Saccenti and Timmerman (2016) created different models with an increasing number of samples. They started with a sample size of 50 (25 control + 25 cases) up to a sample size of 1000 (500 control + 500 cases) and the four experimental data sets, which were used, consisted of nuclear-magnetic-resonance (NMR) spectra of the serum blood metabolites (D.1 & D.2) and urine (D.3 & D.4). The results of the study of Saccenti and Timmerman (2016) showed that the variability decreased with the sample size and sensitivity and specificity increased. These results can also be seen in Figure 16 (sensitivity) and Figure 17 (specificity).

Rienesl et al. (2019) used MIR spectra data for mastitis prediction. They created four different models with different sample sizes but the sample sizes were much bigger than in the study of Saccenti and Timmerman (2016), and started from 1,165 (586 mastitis + 586 healthy) to 3,552 (1,776 mastitis + 1,776 healthy). The result of this study was, that only specificity increased with the number of samples in the calibration set. There were almost no differences concerning to sensitivity.
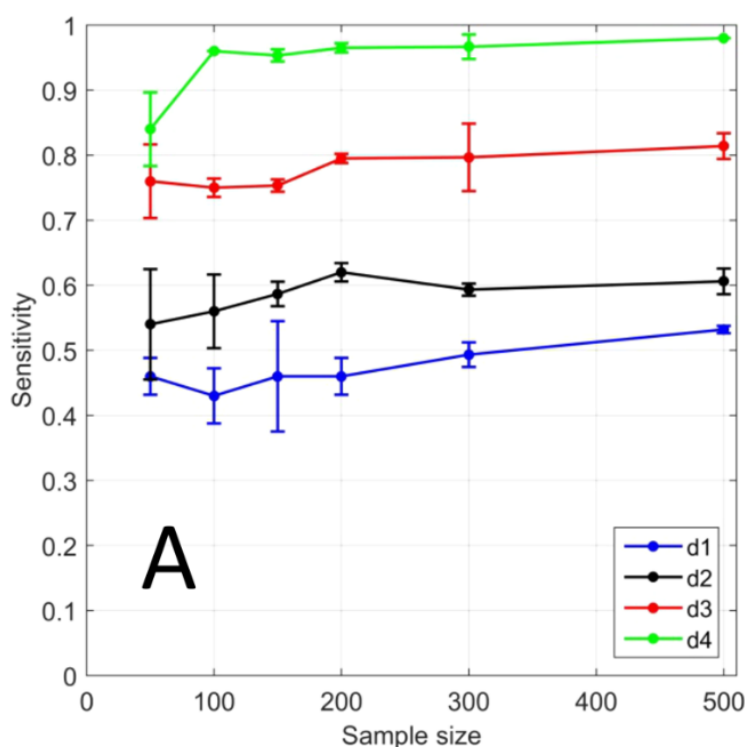

Figure 16: Sensitivity (A) of a PLS-DA model as a function of the total sample size for the discrimination between two group in a case-control design (Saccenti & Timmerman, 2016)
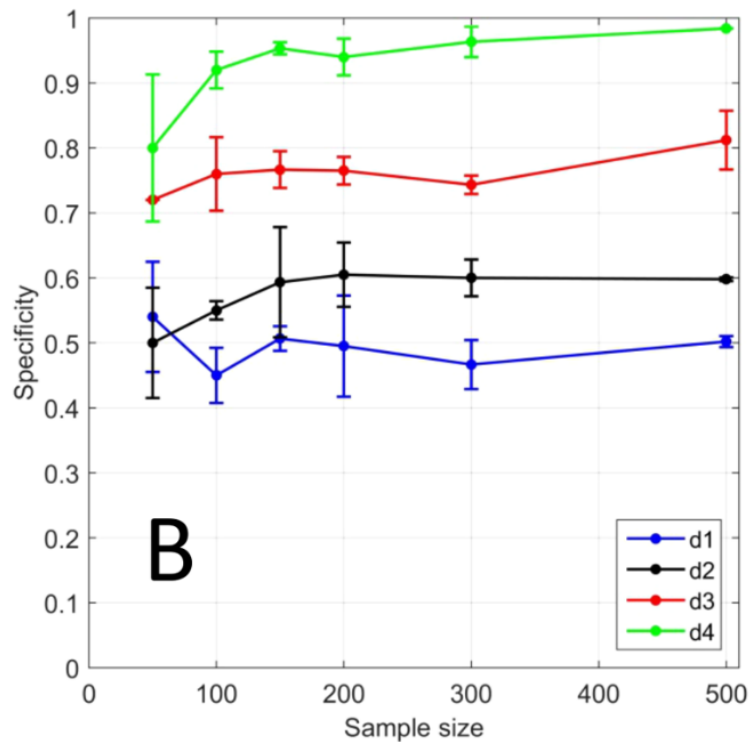
Figure 17: Specificity (B) of a PLS-DA model as a function of the total sample size for the discrimination between two group in a case-control design (Saccenti & Timmerman, 2016)

The sample sizes in the current study were similar to the sample sizes in the study of Saccenti & Timmerman (2016) but there were also smaller and bigger sample sizes for calibration. Sample sizes started with 20 (10 pregnant + 10 open) to 125,162 (62,582 pregnant + 62,582 open). The results showed that the differences of sensitivity (Figure 18) and specificity (Figure 19) between the individual sample sizes were smaller than in the study of Saccenti & Timmerman (2016). In regard to the variability, this study showed also a decrease of the standard deviation with the sample sizes like in the study of Saccenti & Timmerman (2016).

In terms of sensitivity, the results of this study were similar to the results of Rienesl et al. (2019) but there was no regular increase with the sample size for specificity. Yet, more than a few hundred samples per class (open/pregnant) in the calibration set did not substantially improve predictive capacity of the prediction formula applied to the large and imbalanced validation data set.
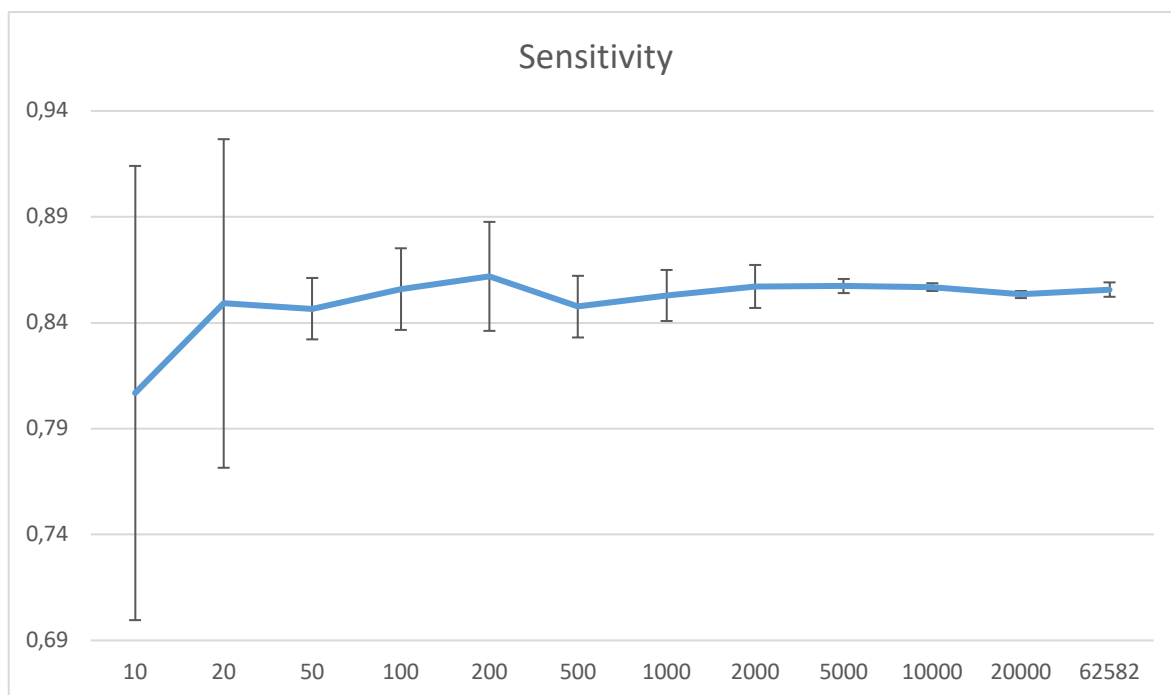
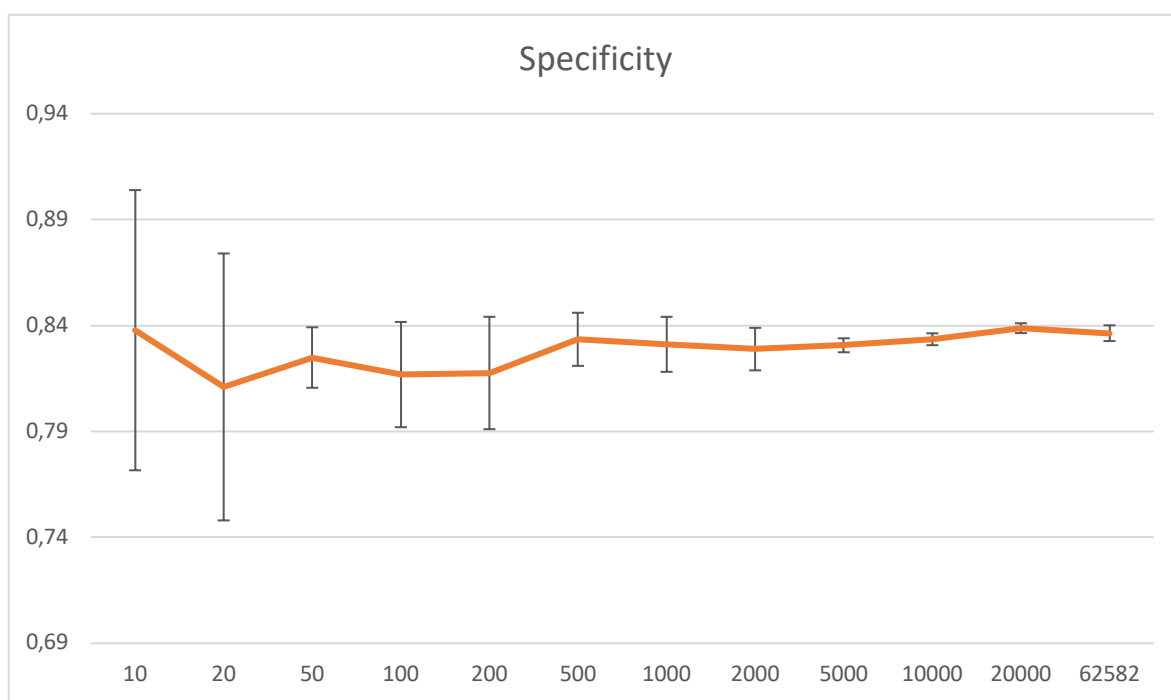*Figure 18: Sensitivities and their standard deviation for different sample sizes in calibration*

Figure 19: Specificities and their standard deviation for different sample sizes in calibration

5.   Conclusion

The main aim of this study was to develop a calibration equation to predict the pregnancy state from routinely recorded MIR spectral data. The results indicate that the prediction of the pregnancy state with MIR spectra data is difficult because of the strong effect of the lactation stage on the MIR spectrum and that fact that cows are typically open in early lactation and pregnant in late lactation. To account for this effect, separate prediction equations were developed for classes of lactation stage and expected pregnancy stage based on the most recent insemination. In general, balanced accuracy was good for cows getting pregnant early in lactation and increased with number of days pregnant. Although the accuracy of prediction is not very high, MIR spectra derived probabilities of pregnancy may be potentially implemented in routine recording in order to provide farmers with an additional tool for fertility management. Whether this is feasible needs to be explored by discussions with farmers.

6.   References

Balhara, A. K., Gupta, M., Singh, S., Mohanty, A. K. and Singh, I. (2013): Early pregnancy diagnosis in bovines: Current status and future directions. *The Scientific World Journal*, Volume 2013, Article ID 958540, 10 pages, http://dx.doi.org/10.1155/2013/958540

Bonfatti, V., Ho, P. N. and Pryce, J. E. (2019): Usefulness of milk mid-infrared spectroscopy for predicting lameness score in dairy cows. *Journal of Dairy Science*, 103(2), in press. https://doi.org/10.3168/jds.2019-17551

Committee on Bovine Reproductive Nomenclature (1972): Recommendations for standardizing bovine reproductive terms. *The Cornell veterinarian,* 62(2), 216-237.

Cowie, T.A. (1948): Pregnancy Diagnosis Tests: A Review. *Commonwealth Agricultural Bureaux Joint Publication*, No.13, 11-17.

D4Dairy (2019). D4Dairy Project. Retrieved from https://d4dairy.com/en/#project. [Accessed 2019 December 30]

Dale, L., and Werner, A. (2017): "MastiMIR"-Ein Mastitis-Frühwarnsystem basierend auf MIR-Spektren. *Paper presented at the conference of DGfZ and GfT, Stuttgard, Germany.* Retrieved from https://www.researchgate.net/publication/320351923

De Marchi, M., Toffanin, M., Cassandro, M. and Penasa, M. (2014): Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. *Journal of Dairy Science*, *97(3*), 1171-1186. https://doi.org/10.3168/jds.2013-6799

Delhez, P., Ho, P. N., Gengler, N., Soyeurt, H. and Pryce, J. E. (2020): Diagnosing the pregnancy status of dairy cows: How useful is milk mid-infrared spectroscopy? *Journal of Dairy Science*, In Press, https://doi.org/10.3168/jds.2019-17473

De Roos, A. P. W., Van Den Bijgaart, H. J. C. M., Hørlyk, J. and De Jong, G. (2007). Screening for subclinical ketosis in dairy cattle by fourier transform infrared spectrometry. *Journal of Dairy Science*, *90*(4), 1761–1766. https://doi.org/10.3168/jds.2006-203

DeWitt, S. F. and Grondin T. M. (2011): Progesteron. In: Clinical Veterinary Advisor: The Horse (Hrsg. Wilson D. A.), Elsevier Inc., US, 956. https://doi.org/10.1016/B978-1-4160-9979-6.00431-1

Egger-Danner, C., Fürst, C., Mayerhofer, M., Rain, C., and Rehling, C. (2018): ZuchtData Jahresbericht 2018. *ZuchtData EDV-Dienstleistungen GmbH, Vienna, Austria*.

Fricke, P. M., Ricci, A., Giordano, J. O. and Carvalho, P. D. (2016): Methods for and implementation of pregnancy diagnosis in dairy cows. *Veterinary Clinics of North America: Food and Animal Practice,* 32(1), 165-180. https://doi.org/10.1016/j.cvfa.2015.09.006

Grelet, C., Fernández Pierna, J. A., Dardenne, P., Baeten, V., and Dehareng, F. (2015): Standardization of milk mid-infrared spectra from a European dairy network. *Journal of Dairy Science*, *98*(4), 2150–2160. https://doi.org/10.3168/jds.2014-8764

Grelet, C., Bastin, C., Gelé, M., Davière, J. B., Johan, M., Werner, A., Reding, R., Fernandez Pierna, J. A., Colinet F. G., Dardenne, P., Gengler, N., Soyeurt H. and Dehareng, F. (2016): Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. *Journal of Dairy Science*, *99*(6), 4816–4825. https://doi.org/10.3168/jds.2015-10477

Hirpa, A., Yehualaw, B., Wube, A., Asnake, A. and Jemberu, A. (2018): Review on Pregnancy Diagnosis in Dairy Cows. *Journal of Reproduction and Infertility*, 9(2), 45–55. https://idosi.org/jri/9(2)18/3.pdf

Ho, P. N., Bonfatti, V., Luke, T. D. W. and Pryce, J. E. (2019): Classifying the fertility of dairy cows using milk mid-infrared spectroscopy. *Journal of Dairy Science*. 102(11), 10460-10470. https://doi.org/10.3168/jds.2019-16412

Humblot (2001): Use of pregnancy specific proteins and progesterone assays to monitor pregnancy and determine the timing, frequencies and sources of embryonic mortality in ruminants. *Theriogenology,* 56(9), 1417-1433. https://doi.org/10.1016/S0093-691X(01)00644-6

ISO (2013): Milk and liquide milk products – Guidelines for the application of mid-infrared spectrometry. ISO 9622 (IDF 141) , International Organization for Standardization, Switzerland.

Klaffenböck, M., Steinwidder, A., Fasching, C., Terler, G., Gruber, L., Mészáros, G., & Sölkner, J. (2017). The use of mid-infrared spectrometry to estimate the ration composition of lactating dairy cows. *Journal of Dairy Science*, *100*(7), 5411–5421. https://doi.org/10.3168/jds.2016-12189

Kuhn, M. (2008): Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*(5), 1–26.

Lainé, A., Bel Mabrouk, H., Dalé, L.-M., Bastin, C. and Gengler, N. (2013): How to use mid-infrared spectral information from milk recording system to detect the pregnancy status of dairy cows. *Communications in agricultural and applied biological sciences*, 79(1), 33-8. http://hdl.handle.net/2268/177706

Lainé, A., Bastin, C., Grelet, C., Hammami, H., Colinet, F. G., Dale, L. M., Gillon, A., Vandenplas, J., Dehareng, F. and Gengler, N. (2017): Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. *Journal of Dairy Science*, *100*(4), 2863–2876. https://doi.org/10.3168/jds.2016-11736

Lantz, B. (2015): Machine Learning with R. Second edition, Packt Publishing Ltd., Page 333ff.

Lawson, B. C., Shahzad, A. H., Dolecheck, K. A., Martel, E. L., Velek, K. A., Ray, D., L., Lawrence, J. C. and Silvia, W. J. (2014): A pregnancy detection assay using milk samples: Evaluation and considerations. *Journal of Dairy Science,* 97(10), 6316-6325. https://doi.org/10.3168/jds.2014-8233

McParland, S., Banos, G., Wall, E., Coffey, M. P., Soyeurt, H., Veerkamp, R. F. and Berry D. P. (2011): The use of mid-infrared spectrometry to predict body energy status of Holstein cows. *Journal of Dairy Science,* 94(7), 3651-3661. https://doi.org/10.3168/jds.2010-3965

McParland, S., Lewis, E., Kennedy, E., Moore, S. G., McCarthy, B., O'Donovan, M., Butler, S. T., Pryce, J. E. and Berry, D. P. (2014): Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. *Journal of Dairy Science*, *97*(9), 5863–5871. https://doi.org/10.3168/jds.2014-8214

Mineur, A., Köck, A., Grelet, C., Gengler, N., Egger-Danner, C., and Sölkner, J. (2017): First results in the use of milk mid-infrared spectra in the detection of lameness in Austrian dairy cows. *Agricultural Conspectus Scientificus*, *82*(2), 163–166. Retrieved from https://www.researchgate.net/publication/325450513

Pohler, K. G., Franco, G. A., Reese, S. T., Dantas, F. G., Ellis, M. D. and Payton, R. R. (2016): Past, present and future of pregnancy detection methods. Applied Reproductive Strategies in Beef Cattle. Des Moines, Iowa, September 7 – 8, 2016.

Olori, V. E., Brotherstone, S., Hill, W.G. and McGuirk (1997): Effect of gestation stage on milk yield and composition in Holstein Friesian Dairy cattle. *Livestock Production Science*, 52, 167-176. https://doi.org/10.1016/S0301-6226(97)00126-7

R Development Core Team (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna.* Retrieved from http://www.R-project.org.

Ricci, A., Carvalho, P.D., Amundson, M. C., Fourdraine, R. H., Vincenti, L. and Fricke, P. M. (2015): Factors associated with pregnancy-associated glycoproteins (PAG) levels in plasma and milk of Holstein cows during early pregnancy and their effect on the accuracy of pregnancy diagnosis. *Journal of Dairy Science*, 98(4), 2502-2514. https://doi.org/10.3168/jds.2014-8974

Rienesl, L. M., Sölkner, J., Khayatzadeh, N. and Köck, A. (2019): Mastitis detection from milk mid-infrared (MIR) spectroscopy in dairy cows. *Master thesis,* University of Natural Resources and Life Sciences, Vienna.

Saccenti, E., and Timmerman, M. E. (2016): Approaches to sample size determination for multivariate data: Applications to PCA and PLS-DA of omics data. *Journal of Proteome Research*, *15*(8), 2379–2393. https://doi.org/10.1021/acs.jproteome.5b01029

Santos, J. E. P., Thatcher, W. W., Chebel, R. C., Cerri, R. L. A. and Galvao, K. N. (2004): The effect of embryonic death rates in cattle on the efficacy of estrus synchronization programs. *Animal Reproduction Rcience,* 82-83, 513-535. https://doi.org/10.1016/j.anireprosci.2004.04.015

SAS Institute Inc. (2017). SAS software 9.4. *SAS Institute Inc., Cary, NC, USA.*

Sinedino, L. D., Lima, F. S., Bisinotto, R. S., Cerri, R. L. A., and Santos, J. E. P. (2014): Effect of early or late resynchronization based on different methods of pregnancy diagnosis on reproductive performance of dairy cows. *Journal of Dairy Science*, 97(8), 4932-4941. https://doi.org/10.3168/jds.2013-7887

Soyeurt, H., Bruwier, D., Romnee, J.-M., Gengler, N., Bertozzi, C., Veselko, D. and Dardenne, P. (2009): Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry. *Journal of Dairy Science*, 92(6), 2444–2454. http://dx.doi.org/10.3168/jds.2008-1734

Soyeurt, H., Dehareng, F., Gengler, N., McParland, S., Wall, E., Berry, D. P., Coffey, M. and Dardenne, P. (2011): Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *Journal of Dairy Science*, *94*(4), 1657–1667. https://doi.org/10.3168/jds.2010-3408

Soyeurt, H., Bastin, C., Colinet, F. G., Arnould, V. M.-R., Berry, D. P., Wall, E., Dehareng, F., Nguyen, H. N., Dardenne, P., Schefers, J., Vandenplas, J., Weigl, K., Coffey, M., Théron, L., Detilleux, J., Reding, E., Gengler, N. and McParland, S. (2012): Mid-infrared prediction of lactoferrin content in bovine milk: potential indicator of mastitis. *Animal*, 6(11), 1830-1838. https://doi.org/10.1017/s1751731112000791

Toledo-Alvarado, H., Vazquez, A. I., de los Campos, G., Tempelman, R. J., Bittante, G. and Cecchinato, A. (2018): Diagnosing pregnancy status using infrared spectra and milk compositions in dairy cows. *Journal of Dairy Science*, 101(3), 2496-2505. https://doi.org/10.3168/jds.2017-13647

Vasconcelos, J. L. M., Silcox, R. W., Lacerda, J. A., Pursley, J. R. and Wiltbank, M. C. (1997): Pregnancy rate, pregnancy losses and response to heat stress after AI at 2 different times from ovulation in dairy cows. *Biology of Reproduction*, 56(1), 230-230.

Vanlierde, A., Vanrobays, M.-L., Dehareng, F., Froidmont, E., Soyeurt, H., McParland, S., Lewis, E., Deighton, M. H., Grandl, F., Kreuzer, M., Gredler, B., Dardenne, P. and Gengler, N. (2015:. Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra. *Journal of Dairy Science*, *98*(8), 5740–5747. https://doi.org/10.3168/jds.2014-8436.