Detecting dairy cows' lying behavior using noisy 3D ultra-wide band positioning data

February 17, 2022

I. Adriaens[1,*], W. Ouweltjes[1], M. Pastell[2], E. Ellen[1], C. Kamphuis[1]

[1] Animal Breeding and Genomics, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands.
[2] Luke, PLF group, Production Systems, Natural Resources Institute Finland (Luke), Latokartanonkaari 9, 00790 Helsinki, Finland.
[*] *corresponding author*: ines.adriaens@wur.nl

1 Abstract

In precision livestock farming, technology-based solutions are used to monitor and manage livestock and support decisions based on on-farm available data. In this study, we developed a methodology to monitor the lying behavior of dairy cows using noisy spatial positioning data, thereby combining time-series segmentation based on statistical changepoints and a machine-learning classification algorithm using bagged decision trees. Position data (*x, y, z*-coordinates) collected with an ulta-wide band positioning system from 30 dairy cows housed in a freestall barn were used. After the data preprocessing and selection, statistical changepoints were detected per cow-day (no. included = 331) in normalized 'distance from the center' and (z) time series. Accelerometer-based lying bout data were used as a practical ground truth. For the segmentation, changepoint detection was compared with getting-up or lying-down events as indicated by the accelerometers. For the classification of segments into lying or non-lying behavior, two data splitting techniques resulting in 2 different training and test sets were implemented to train and evaluate performance: one based on the data collection day and one based on cow identity. In 85.5% of the lying-down or getting-up events a changepoint was detected in a window of 5 minutes. Of the events where no detection had taken place, 86.2% could be associated with either missing data (large gaps) or a very short lying or non-lying bout. Overall classification and lying behavior prediction performance was above 91% in both independent test sets, with a very high consistency across cow-days. This resulted in sufficient accuracy for automated quantification of lying behavior in dairy cows, for example for health or welfare monitoring purposes.

Keywords: *spatial data; ultra-wide band technology; dairy cow; lying behavior*

2 Introduction

Precision livestock farming solutions typically aim at supporting monitoring and decision taking by farmers using on-farm sensors measuring animal behavior, performance and production [1]. The raw data used to generate decision support are often noisy time series, prone to errors and variation caused not only by sensor failure or the harsh and changing farm environments in which they operate, but also by the animals' specific physiology itself. The resulting complexity and magnitude of the raw data render them hard to interpret as such by farmers or other end-users. Consequently, these data have little value without proper (pre-)processing algorithms that translate the raw measures in information informative for the targeted end-users.

In dairy production, precision technologies are vastly deployed and implemented [2, 3]. The reason for the dairy sector being pacesetter in this area, is groups of animals are typically much less homogeneous compared to other livestock species and therefore management at group level is less applicable. Additionally, dairy cows are highly valuable but rather vulnerable, rendering individual monitoring crucial to optimize production, welfare and sustainability. Because of the physiological stress these animals endure during lactation, timely and specific interventions obviate animal suffering and financial losses. As modern dairy farms grew larger over the past decade, investments in sensor technology to guide these interventions became increasingly justifiable [4]. Out of the many technologies available, a system monitoring cow position and its derived behavioral features not only promises to disclose cow health, but might also reveal welfare and social interactions - aspects that become increasingly important in the livestock production landscape. Today's commercialized positioning systems, however, mainly serve to locate cows for e.g., treatment or when they don't go milking. Monitoring specific cow behaviors offers new paths both for research and commercial decision support systems that can help the farmer manage their herd, optimize production and quickly act upon disease or welfare problems. A continuous and essential step to better unlock the potential of cow behavioral analyses is the development of new ways to process data from sensor technologies that allow precise and timely interpretation and extraction of actionable information [5]. As such, extra value can be created from existing technology.

Lying behavior has been shown to change upon a changing health and welfare status [6]. For example, lameness will lower the number of times an animal gets up or lies down and increases general lying bout duration. Similarly, udder infections in which an animal becomes very sick, or metabolic problems affecting rumination time, will alter the lying behavior [7]. Accurate detection and monitoring over time of lying thus has potential to reveal health and welfare status, contribute to new precision phenotypes, and evaluate e.g., housing situations or management practices in an accurate and non-invasive manner. One technology to do so is via 3-dimensional spatial data, such as provided via modern ultra-wide band (**uwb**) positioning systems currently being developed and commercialized.

Ultra-wide band technology allows the transmission of high amounts of data over small distances with very low energy. In an indoor positioning system, Radio-Frequency identification signals are transmitted across a wide bandwidth and captured by an antenna. The tags worn by the individual cows allow precise and frequent localization of the animals with low power usage, even in cluttered indoor environments [8]. Upon development of appropriate data interpretation algorithms, indoor positioning systems allow studying and monitoring cow behavior, including general activity, resting, feeding, drinking and social interactions with a single sensor system, giving it a relative advantage over e.g. commercially available accelerometer systems. Similarly, video-based systems have the challenge of cow-identification, sufficient spatial covering, and high computation power requirements. Despite its continuous development and high potential for animal monitoring, uwb-based positioning is yet sparingly adopted for livestock applications. As for many new sensor technologies, the main reason for this is the lack of algorithms that translate raw data into information valuable to the farmer [9]. In case of indoor positioning systems, data interpretation is complicated by the inaccuracy and noise in the time series, missing data, and its (unpredictable) heteroscedasticity [10, 11]. The latter partly results from differences in behavior, but previous research also highlighted dependency on the position of the animal in the barn with regard to the antenna and interactions of the signal with metal (e.g. the feeding rack) and water bodies (e.g., other cows). These aspects hinder straightforward interpretation of the positioning data and its derivatives (e.g., distance traveled), also preventing wider adoption. Nonetheless, as dedicated processing of these data would tremendously increase data interpretation potential, for example for the classification of behavior, several studies on this topic have been published in the past few years [12, 13, 14, 15].

There is a high need for new methods that elegantly integrate and interpretat on-farm collected longitudinal data on which decision support can be based. Additionally, automated, continuous and non-invasive detection of lying behavior for health and welfare monitoring based on spatial data has not been described in the past. In this study, a two-step methodology to identify lying behavior of dairy cows using a uwb-based indoor positioning system was developed and validated against the lying bouts returned by a commercial accelerometer-based system. The methodology relies on segmentation via the detection of simultaneous changepoints in two position-derived time series, considering both the *(x,y)* and in the (z)-direction, while avoiding fixed thresholds or severe assumptions on the statistical properties of the data. The individual segments are in a second step classified as 'lying' or 'non-lying' based on a set of statistical properties.

3 Materials & methods

3.1 Data collection

Data were collected at the Dairy Campus research facilities of Wageningen University and Research in Leeuwarden, the Netherlands, during two periods of five days in two successive weeks in 2019 (July 3 to 8 and July 10 to 15, both periods with normal weather conditions with temperatures between 10 and 20°C). Two groups of cows, one housed in a freestall barn with a straw deep litter bedding and one in a freestall with synthetic flooring, were equipped with uwb-positioning tags on the upside of a neck collar (Ubisense, Cambridge, UK and Noldus, Wageningen, the Netherlands) and accelerometers attached to right hind leg (IceQube® pedometers, IceRobotics, Edinburgh, United Kingdom). It is important to note that the Ubisense technology relies on different methods to determine (*x,y*)-position compared to (*x*)-position. The first is calculated based on *time difference of arrival*, whereas the latter is derived from the *axis of arrival*, which makes the (*z*) more dependent on e.g., orientation of the tags. Each group consisted of 16 cows selected based on production level, age and lactation stage such that the characteristics were comparable across each group. The cows were milked twice daily in a rotary parlor and fed *ad libitum* with a partial mixed ration complemented with concentrates individually rationed based on production level.

3.2 Lying behavior

As continuous visual observation of the animals' behavior is too laborious over a longer period of time, the lying bouts returned by the IceQube accelerometers were used as the benchmark 'ground truth' for lying behavior. Despite this is a sensor-based measure and not visual observation which would be the true gold standard, it allows to include multiple cows simultaneously, with minimal labor and for a longer period of time, and it has been shown to have sufficient accuracy to detect the actual lying behavior, with r $>$ 0.99 **[13]**. For each cow, the timestamp of each lying down or getting up event was retrieved from the IceQube software. These data were visually assessed to verify time synchronization and cow identity across the different sensor systems. Only data for which in that time period both uwb and IceQube data were available were retained. More specifically, for each cow, data were kept from the first available IceQube lying bout onward until the end of the last lying bout registered, such that the analysis was carried out on the data for which accelerometers were certainly attached to the animals. This prevented that a lack of lying bout registrations was not caused by cows not wearing a sensor. Two out of the 32 cows were excluded from the study because no ground truth lying bouts were registered due to a technical problem with the IceQube sensors.

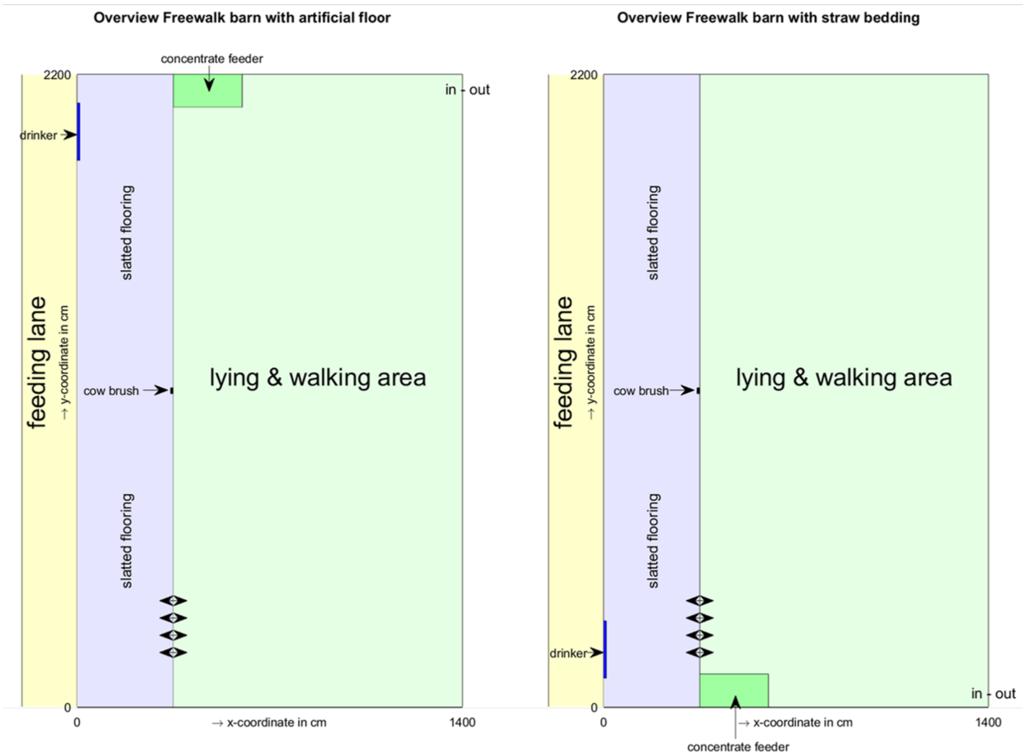
3.3 Ultra-wide band data editing

Raw binary data were extracted from daily Tracklab back-up files (.tlp) (Noldus, Wageningen, the Netherlands) and converted with Python 3.7 into *(x,y,z)*-position time series containing one measurement per second per cow. All further data processing was done using Matlab 2018b and 2020b (The MathWorks Inc., Natick, Massachusetts, USA). The *(x,y,z)*-position was expressed relative to a pre-specified origin *(x,y,z)=(0,0,0)*. In the barns at Dairy Campus, the *(x)*- coordinate gives the position in the direction of the feeding racks (range 0 to 23m in the first barn and 23 to 46m in the second barn), whereas the *(y)*-coordinate represents the position perpendicular to the feeding alley (range 0 to 14m). A plan of the barn is shown in **figure 1**. The *(z)*-position can be considered the height of the tag on the neck collar. When the y was larger than 11.5m, the animals were in the slatted flooring (feeding) area, in which it was considered they did not lie down (as formally confirmed by the IceQube data). To interpret the raw position time series and derive cow behavior from them, multiple data editing steps were implemented to deal with noise and missing data (missing data = on average 43% per day, small gaps and absent data due to milking included). First, outliers indicating a position outside the barn edges were replaced with the edge value when it were single measurements likely caused by normal measurement inaccuracy. When multiple successive measurements were registered out of the barn edges, they probably resulted from a lost tag that was put aside by the animal caretakers (in our dataset, this happened during 11 cow-days). These measurements were replaced by missing values. Second, based on a data exploration step (not further detailed in this paper), a methodology to manage missing data was developed and implemented. How we dealt with the missing data depended on (1) the gap size and (2) the amount of non-missing data in predefined window preceding the gap. Missing data always occurred at cow-measurement level, i.e., if data were unavailable, both the *(x,y)*- and (z)-position lacked. For gaps smaller than 60 seconds, we assumed that the cow's behavior would remain constant, or the error made when this assumption was untrue would be negligible. In this case, the missing data were imputed by sampling them from a normal distribution with mean and standard deviation calculated from the data preceding the gap in a window of twice the gap size in each dimension. For gaps between 60 and 180 seconds, making assumptions on the consistency of the behavior was more tricky but these gaps could still be due to failure of the sensor system or interference with the barn environment. For these gaps, we used a simple linear interpolation with added noise based on the average standard deviation of the data. Missing data in gaps longer than 180 seconds were left without data, as these often resulted from the animals not being in the barn e.g. during milking. Assumptions on these longer lasting gaps could not be made and were not of interest for this study, as in these cases cows are not expected to lie down. A third data editing step consisted in smoothing the (x)-, (y)- and (z)-data with a moving median filter in a window of 45 seconds. In order to make sensible assumptions for the settings of the changepoint analysis, data of each cow-day were analyzed separately (i.e., a separate segmentation was implemented per cow-day time series).

3.4 Changepoint analysis for segmentation

Changepoints are time instants or samples in which the statistical properties (i.e. statistical distribution) of a (time) series abruptly change. In this study, we detected and combined the individual changepoints per cow per day in two time series of (x,y,z)-coordinate positioning data. Intuitively, one could argue to mainly rely on the position in the vertical (z) direction (height), as a cow that lies down is expected to remain in a lower and more stable position compared to when she is not lying down. However, the (z)-position was found (unpublished data exploration step) to be the most unreliable and noisy (range, variability,...) of all three coordinates. Its inaccuracy was variable in time and space, and depended on e.g., the position in the barn, the behavior and speed of the animals, the collar attachment, the calibration settings and individual interactions between tags. Similarly, relying on detection of a relatively stable position in the (x,y)-direction (which is unmistakably true during lying bouts) is imprecise and insufficient for lying behavior detection as well, as cow activity varies over the day, and oftentimes animals stand still for a longer period of time apart from their lying bouts, for example when grooming other animals, feeding, drinking or ruminating. These periods of 'standing' inactivity might additionally depend on accessibility lying places, hierarchy, climate of the barn, etc. In this study, we chose to work on a combination of two position-derived

Figure 1: Plans of the barns in which the data were recorded



time series. The first is the (z)-coordinate (height) of the animals, as this is the most straightforward one. The second time series is the 'center distance' (**CD**), i.e. the position relative to the center of the barn. The main advantages of using CD and not the raw (x,y)-position is that it summarizes position and movement of the animals in a single signal, is less dependent on the actual direction of movement, and has a lower variability and range. Should a cow move in a perfect circle around the center of the barn, however, CD remains constant (as is the case when a cow stands still or lies down). We assumed that this would be extremely rare, and when it would happen for a short period of time, this would not impair the analysis because movement as such causes the signal to be more variable, which also changes the statistical properties of the time series. Before the segmentation, the CD and (z) time series were normalized with a min-max standardization per cow over the entire dataset as follows:

$$x_{i,norm} = \left[\frac{x_i - min(x)}{max(x) - min(x)}\right]$$

with x_i the z or CD values at time i.

The changepoint analysis relies on a parametric method that partitions both time series simultaneously in K segments based on the minimization of the following cost function *J(K)*:

$$J(K) = \sum_{r=0}^{K-1} \sum_{i=k_r}^{k_{r+1}-1} \Delta(x_i; \chi([x_{k_r} \cdots x_{k_{r+1}-1}]))$$

with

$$\sum_{i=k_r}^{k_{r+1}-1} \Delta(x_i; \chi([x_{k_r} \cdots x_{k_{r+1}-1}])) = ((k_{r+1}-1)-k_r+1)*log(var([x_{k_{r+1}-1} \cdots x_{k_r}]))$$

and

$$var([x_{k_{r+1}-1} \cdots x_{k_r}]) = \frac{1}{(k_{r+1}-1)-k_r+1} \sum_{i=k_{r+1}-1}^{k_r} (x_i - mean([x_{k_{r+1}-1} \cdots x_{k_r}]))$$

in which K is the number of changepoints, dividing the time series in *K+1* segments, β is the penalty function, here restrained such that at most 60 changepoints are found per cow-day, because otherwise the number of changepoints would equal the number of data points as this minimizes the total cost. As adding changepoints in general lowers the cost function, it is normal that the number of changepoints found is equal to the maximum set beforehand. Because the variability in the data was high and thereby unpredictable, a mathematical penalty function for restricting the number of changepoints detected could not be found. x_{k_r} is the $r^t h$ z or CD value in segment *k*. Besides in 'number', also a restriction was set to the minimum distance between two changepoints: they needed to be at least 300 measurements apart (i.e. the lying or non-lying duration was at least 5 minutes). Other data-based algorithms (i.e., using variability and expected minimal cost reduction) have been explored, but because of the heteroscedastic nature of the data, could not be used for this study. The changepoint search algorithm used is based on a pruned exact linear time algorithm using dynamic programming, as proposed by Killick et al. **[16]**, having the advantage that it is mathematically exact and has a linear computational cost with the number of data points.

3.5 Data split

To evaluate the performance of the classification algorithm, its performance was evaluated using two different data splits, one based on time and one based on cow identity. For both, we chose to use a smaller portion of the data for training than for testing (approximately 33-66%), unlike what is usual in machine learning practices. However, we preferred this data split as (1) the method described here is very robust, so a minimal amount of training data sufficed to achieve accurate predictions and adding more data did not improve the accuracy, and (2) this situation mimics an on-farm situation where little training data is available. The first data split (alike the more classical machine learning approach) uses data from 10 randomly chosen cows (33%) for the model training, and 20 animals (66%) as the independent test set. The second approach corresponds to a situation on farm in which current and historical data are used for training and the algorithm needs to perform well in a future situation. Here, data of the 3 first days of the dataset were assigned to training set, after which classification performance was evaluated on the remaining 9 days of data. One cow's data only started at day 4, and was therefore not included in this training set as the animal would not have been present in the training period.

3.6 Segment classification

To move from segments to lying behavior, we classified each segment as 'lying' or 'non-lying' based on its (statistical) properties, including the level and variability for the normalized data, a categorical variable to indicate whether the cow was in the slatted flooring area, the length of the segment, the number of outliers, the gapsize, and the segment range. An overview of these features is given in **table 4**. The classification was done using a 'bagged' (i.e., bootstrap-aggregated) tree algorithm which consistently performed best on our data independently of input data and split. As opposed to individual decision trees (which tend to over fit), bagged trees combine (i.e., use an ensemble) the results of many trees, improving generalization. The algorithm uses a random subset of predictors at each decision split (similar to random forest classification) and minimizes the classification error at each split. The model was trained with 5-fold cross-validation to determine the optimal hyper parameters for the number of learning cycles (i.e., 30) and trees. For the bootstrapping, each time one segment was sampled with replacement to grow a new tree. As in some cases a 'true' change happened within a segment, a threshold of 50% was applied to calculate the binary outcome variable: if more the 50% of the segment's data corresponded to a lying bout, it's ground truth was taken as 'lying' and vice versa. The features were selected such that there was no multicollinearity across them.

3.7 Performance evaluation

Two aspects of the methodology are important to achieve a good performance: (1) the segmentation accuracy, i.e. are the true changes from lying to non-lying and vice versa accurately detected; and (2) the classification performance in terms of accuracy per segment and corresponding total lying duration per cow-day. For the first, we calculated how many of the true changes have a changepoint associated with them within a window of 5 minutes. Given the length of the lying bouts, this is considered as an acceptable margin for detection. When no detected changepoint was associated with the true change, we assessed potential causes, including e.g., missing data. The second was assessed using the confusion matrix comparing true and false classifications and the total accuracy, for the entire dataset as well as at cow and at cow-day level. We additionally compared the total lying down duration per cow-day in a similar way.

4 Results

4.1 Data overview

A total of 30 cows, with each having between 4 and 12 days of data available were included in the study. These cows had parities between 1 and 7, and were on average 188 (range 119 to 243) days in lactation. An overview of the cow characteristics is given in **table 1**.

Table 1: Overview of cow characteristics

Name	average	std	min	max
Parity	2.77	1.50	1.00	7.00
Lactation stage	188.16	43.49	119.00	243
Daily milk yield	26.95	6.01	12.68	41
Fat%	4.72	0.45	4.01	5.44
Protein%	3.38	0.23	2.94	4.06
Lactose%	4.49	0.11	4.23	4.68
SCC*1000c/mL	200.08	212.05	24.75	1035

Over the measurement period, in total 2720 lying bouts were detected with the IceQube sensors. From these, 97 bouts were shorter than 10 minutes. Per cow, an average number of 90.6±24.4 lying bouts per cow were included, with an average duration of 85.3±19.8 minutes per bout across cows. Cows had on average 8.2±1.8 lying bouts per day (range: 4.5 to 11.3) and spent 8.23 hours lying down in total. The within-bout level and standard deviation of the z time series, and the standard deviation of the CD across lying and non-lying bouts are given in **table 2**. From this, it is clear that statistical properties of the chosen time series differ across lying and non-lying behavior, which is the basis of our analysis.

Table 2: Statistical properties of the time-series data across lying and non-lying bouts

	lying				non-lying			
	average	**std**	**min**	**max**	**average**	**std**	**min**	**max**
average z	0.71	0.10	0.49	0.89	1.21	0.09	1.06	1.34
std z	0.25	0.05	0.14	0.33	0.32	0.03	0.27	0.40
average znorm	0.28	0.04	0.20	0.36	0.48	0.04	0.42	0.53
std znorm	0.10	0.02	0.06	0.13	0.13	0.01	0.11	0.16
std CD	0.45	0.10	0.29	0.73	1.68	0.23	1.23	2.18
std CDnorm	0.04	0.01	0.02	0.06	0.13	0.02	0.10	0.17

4.2 Changepoint detection

Of all 5443 ground truth changes in the dataset, 85.5% had a changepoint detected within 5 minutes. Per cow-day, this corresponds to 2.3 changes not identified accurately with the changepoint analysis. From these unidentified changes, 50.3% were linked to changes at a moment that there were more than 15 minutes of missing values in the surrounding hour, and 62.2% of these 50.3% were in a segment with at least 20% missing data. Additionally, 23.9% of these false negatives were within less than 20 minutes from another ground truth change, and thus associated with a very short segment length (**table 3**). At cow level, the performance remained more or less constant, with 14.2% of the changes not detected within 5 minutes of the ground truth and up to 93% associated with missing data. It is expected that part of the changes not being correctly identified with the changepoint analysis is also due to the ground truth not being perfect but this can, with the current dataset, not be verified.

Table 3: Overview of correctly and incorrectly detected changepoints corresponding to lying down or getting up

	No.	%
Ground truth changes	5443	100
Detected changepoints within 5 minutes of ground truth	4654	85.5
Not detected changepoints within 5 minutes of ground truth	789	14.5
with >15' missing values in surrounding hour	397	50.3
with previous/next changepoint within 20'	189	23.9

4.3 Classification performance for cow identity-based data split

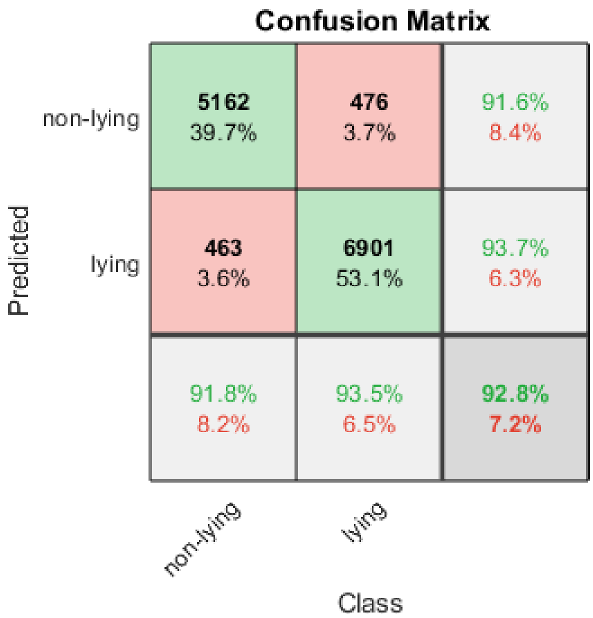
The first split was based on cow identity, and the training dataset consisted of 7024 segments (35%) from 10 animals, from which 3206 segments represented non-lying behavior (45.64%). The independent test set contained 13002 segments. The cross-validation accuracy on the training dataset was 91.7%, and the overall prediction accuracy of the test set was 92.8%. The confusion matrix is shown in **figure 2**. In total, the test set contains 5625 non-lying segments, from which 5162 were correctly classified, rendering a non-lying classification accuracy of 91.8%. From the 7377 lying segments in the test set, 6901 were correctly classified, corresponding to a classification accuracy of 93.5% for the lying behavior. In terms of lying duration, the total predicted non-lying time was 2480h, being 115h different from the ground truth non-lying time of 2595h (percent deviation = 4.4%). The total lying time was estimated as 2327h, which is 141h less than the actual lying time of 2468h in the test set (difference 5.7%).

Per cow-day, the average classification accuracy at the segment level was 92.8% with a minimum accuracy of 78.7% and a maximum accuracy of 100% (**figure 3**, left panel). This corresponded to an average error of 7.1% in the estimation of lying duration at cow-day level (**figure 3**, right panel).

4.4 Classification performance for time-based data split

In the second split based on time, 5138 segments were included in the training dataset of day 0,1 and 2, from 29 cows. The confusion matrix is shown in **figure 4**. In the training set, 2229 (i.e. 43.4%) segments represented 'non-lying' behavior. The test set contained 14888 segments from 30 cows. The cross-validation accuracy on the training set was 92.3%. In the test set, 6102 out of 6602 segments were correctly classified as non-lying (accuracy 92.4%), whereas 7634 out of 8286 segments were correctly classified as lying (accuracy 92.1%). The total predicted non-lying duration over the entire dataset was 2853h, whereas the ground truth was a non-lying duration of 2980h, giving a difference of 127h (4.27% over the entire test set). The predicted and ground truth lying duration in the test set were 2612h and 2830h respectively, corresponding to a deviation of 217h or 7.7%.

Figure 2: Confusion matrix for the split based on cow identity

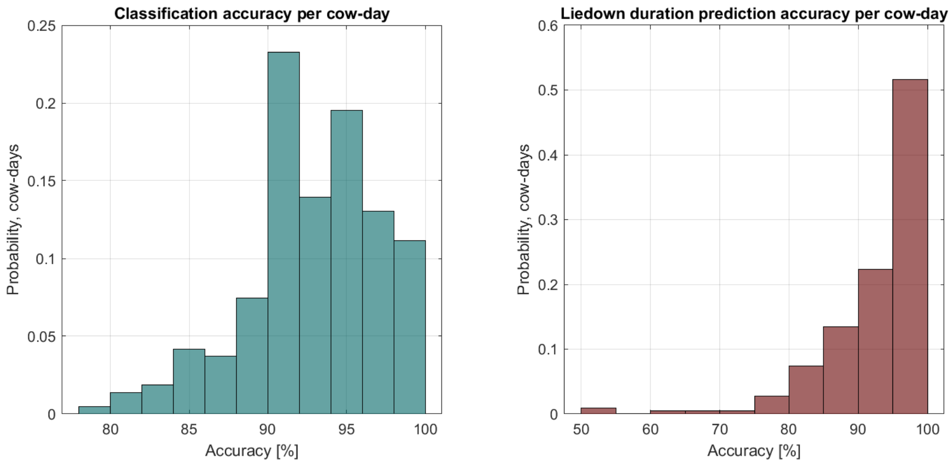


Per cow-day, the average classification accuracy at the segment level was 92.3% with a minimum accuracy of 78.3% and a maximum accuracy of 100% (**figure 5**, left panel). This corresponded to an average error of 7.8% in the estimation of lying duration at cow-day level (**figure 5**, right panel).

5 Discussion

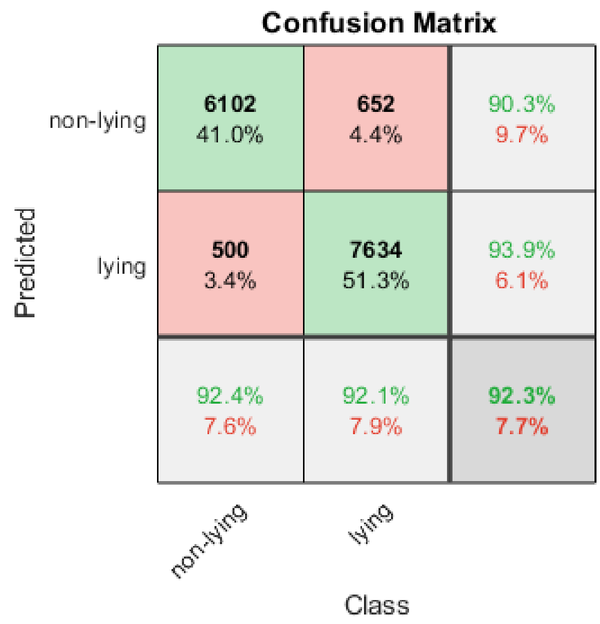
In this study, a methodology was developed to distinguish lying from non-lying behavior of dairy cows based on spatial uwb (x,y,z)-positioning data in a freestall barn, combining a segmentation and classification step. A high segmentation performance overall was reached, with many of the true changes indeed resulting in an alteration of statistical properties and corresponding changepoint in the selected time series. Previous (unpublished) results showed that a combination of time series, and finding simultaneous changepoints was necessary to achieve good results, which supports the general idea that more data integration is needed to achieve good performance in on farm situations in which data are often noisy and prone to many kinds of errors. This was confirmed by the fact that mainly data-quality issues related to missing data and atypical lying behavior (i.e. short lying and non-lying bouts) prevented reaching a higher performance in the segmentation step. The overall and at cow-day level classification performance was high, with accuracies above 91% independent of data split, demonstrating that our methodology is robust and has high practical value. We evaluated the performance of the methodology based on a data split that contained most data in the independent test set and not in the training set

Figure 3: Prediction accuracy at cow-day level for the split based on cow identity



to mimic practical on-farm situation. Robustness of the algorithm is demonstrated by the fact that both the cow identity-based split and the time-based split performed equally well. Future research can investigate the performance of the model when using different position-measuring technologies or in other farm settings and over a longer period of time.
By cross-comparing sensor-based predictions instead of using visual observation, we could validate the methodology with quite an extensive dataset in contrast to what is usual when visual observations are used (e.g.,**[17]**). For example Kok et al. **[18]** used a similar approach for validation of the IceQube accelerometers for lying behavior, comparing the prediction results of two sensors attached to the same cow. Working with spatial data has proven challenging, and e.g., attempts to implement data-based penalty functions for restricting the number of changepoints, failed. This is mainly due to the enormous heteroscedasticity in these data, which depends on multiple factors such as the cow, the time of the day, the behavior, factors interfering with the sensor system, etc., for which we cannot account mathematically. Applying purely black-box approaches generally results in insufficient robustness, interpretability and generalisability **[19, 20, 21]**. Therefore, introducing expert knowledge in animal monitoring algorithms, for example for the data-preprocessing steps, remains essential to make them useful for the end-users. In the current study, expert knowledge was used to pre-process and impute the data, to decide how to combine the spatial data into time series of interest for lying behavior and set the number and distance of changepoints. Other algorithms have been developed to automatically detect lying behavior in dairy cows, for example using machine vision solutions **[12]**. The latter study reported a high sensitivity of 92% as well, but this was not based on lying duration, but on whether there were or weren't animals lying in a cubicle in a specific frame, ignoring the longitudinal importance of the data and restricting its current applicability on farm. Additionally, our algorithm was developed in a freestall barn without cubicles. In cubicle barns, position of the cows in the lying places could be considered as a variable as well, which allows tailoring the

Figure 4: Confusion matrix for the split based on time



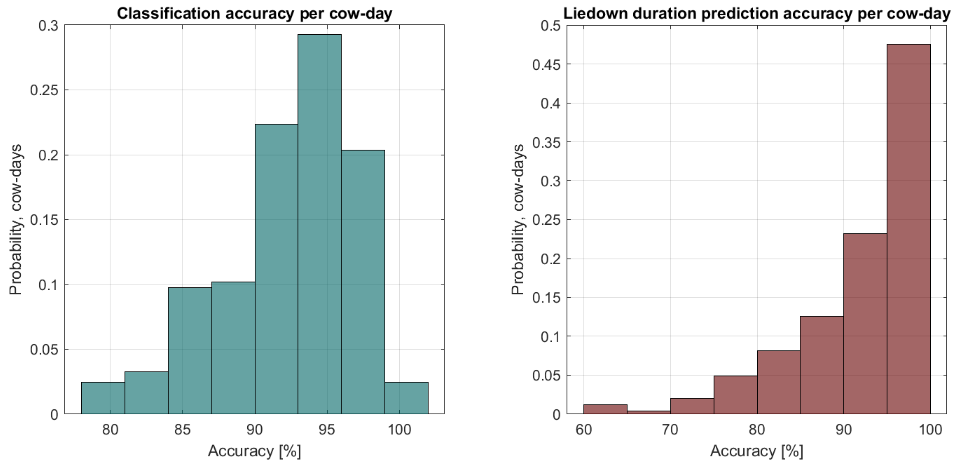
algorithm to different barn circumstances.

In this study, we demonstrated how correct processing of aspecific positioning data (i.e., the system is not designed as such for lying behavior only) allows to use one system for multiple purposes, maximizing the value of a single investment. In a practical setting, the developed methodology shows sufficient performance for monitoring lying behavior of dairy cows over time. For example, the algorithm could be used to create time-series data of lying behavior (duration, bout length), which can be assessed with additional interpretation tools such as individual control charts **[22, 23]**. Combining these at group or at herd level, for example into time budgets allocated to certain behaviors of interest, can also indicate cow health and welfare dynamics of the animals **[6]** and allows automated monitoring with little manual labor. We believe that our methodology can be generalized to other sensor data sources as well.

6 Conclusions

In this study, we developed a methodology to predict certain aspects of the lying behavior of dairy cows from spatial data with the use of time-series segmentation and a subsequent classification algorithm. The methodology relies on differences in statistical properties across the behavior of interest. The overall performance, both when considering a cow-based and a time-based data split to train and evaluate the methodology, was above 92%. Missing data pose the main challenge to reach even higher accuracies, but this doesn't necessarily impair the interpretation of the current results and usability of the method in a

Figure 5: Prediction accuracy at cow-day level for the split based on time



practical setting. Generalization of the segmentation-classification method to other behaviors and other sensors was identified as a potential route to improve on-farm data interpretation for decision support.

7 Acknowledgements

This work was conducted as part of a study sponsored by the Dutch Ministry of Economic Affairs (TKI Agri & Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin, and was also part of the framework for Policy Support with Research theme “Data driven & High Tech” (Sensing potential, KB-38-001-008 and Artificial Intelligence KB-38-008-002) funded by the Dutch Ministry of Agriculture, Nature and Food Quality. Data collection was done in the context of the ERA-NET SusAn "FreeWalk" project, financially supported by the European Union’s Horizon 2020 Research and Innovation Program under grant agreement No. 696231. We thank the Livestock Technology data team of KU Leuven (Department of Biosystems, Belgium) for their valuable inputs and fruitful discussions on the methodology.

References

[1] T. M. Banhazi, H. Lehr, J.L. Black, H. Crabtree, P. Schofield, M. Tscharke, and D. Berckmans. Precision livestock farming: an international review of scientific and commercial aspects. *International Journal of Agricultural and Biological Engineering*, 5(3):1–9, 2012.

[2] D. Lovarelli, J. Bacenetti, and M. Guarino. A review on dairy cattle farming: Is precision livestock farming the compromise for an environmental, economic and social sustainable production? *Journal of Cleaner Production*, 262:121409, 2020.

[3] A.H. Stygar, Y. Gómez, G.V. Berteselli, E. Dalla Costa, E. Canali, J.K. Niemi, P. Llonch, and M. Pastell. A systematic review on commercially available and validated sensor technologies for welfare assessment of dairy cattle. *Frontiers in Veterinary Science*, 8:177, 2021.

[4] M.R. Borchers and J.M. Bewley. An assessment of producer precision dairy farming technology use, prepurchase considerations, and usefulness. *Journal of dairy science*, 98(6):4198–4205, 2015.

[5] E.A. Eckelkamp and J.M. Bewley. On-farm use of disease alerts generated by precision dairy technology. *Journal of Dairy Science*, 103(2):1566–1582, 2020.

[6] C. B. Tucker, M. B. Jensen, AM. de Passillé, L. Hänninen, and J. Rushen. Invited review: Lying time and the welfare of dairy cows. *Journal of Dairy Science*, 104(1):20–46, 2021.

[7] J.M. Piñeiro, B.T. Menichetti, A.A. Barragan, A.E. Relling, W.P. Weiss, S. Bas, and G.M. Schuenemann. Associations of pre-and postpartum lying time with metabolic, inflammation, and health status of lactating dairy cows. *Journal of dairy science*, 102(4):3348–3361, 2019.

[8] Y. Zhou, C.L. Law, and J. Xia. Ultra low-power uwb-rfid system for precise location-aware applications. In *2012 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 154–158. IEEE, 2012.

[9] R. García, J. Aguilar, M. Toro, A. Pinto, and P. Rodríguez. A systematic literature review on the use of machine learning in precision livestock farming. *Computers and Electronics in Agriculture*, 179:105826, 2020.

[10] K. Ren, P.P. Nielsen, M. Alam, and L. Rönnegård. Where do we find missing data in a commercial real-time location system? evidence from 2 dairy farms. *JDS Communications*, 2(6):345–350, 2021.

[11] Matti Pastell, Lilli Frondelius, Mikko Järvinen, and Juha Backman. Filtering methods to improve the accuracy of indoor positioning data for dairy cows. *Biosystems Engineering*, 169:22–31, 2018.

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

[13] M. Borchers, Y.M. Chang, I.C. Tsai, B.A. Wadsworth, and J.M. Bewley. A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors. *Journal of Dairy Science*, 99(9):7458–7466, 2016.

[14] J. Maselyne, M. Pastell, P.T. Thomsen, V.M. Thorup, L. Hänninen, J. Vangeyte, A. Van Nuffel, and L. Munksgaard. Daily lying time, motion index and step frequency in dairy cows change throughout lactation. *Research in veterinary science*, 110:1–3, 2017.

[15] S.J. Hendriks, C.V.C. Phyn, J.M. Huzzey, K.R. Mueller, S.A. Turner, D.J. Donaghy, and J.R. Roche. Graduate student literature review: evaluating the appropriate use of wearable accelerometers in research to monitor lying behaviors of dairy cows. *Journal of Dairy Science*, 103(12):12140–12157, 2020.

[16] R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012.

[17] J.A. Vázquez Diosdado, Z.E. Barker, H.R. Hodges, J.R. Amory, D.P. Croft, N.J. Bell, and E.A. Codling. Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. *Animal Biotelemetry*, 3(1):1–14, 2015.

[18] A. Kok, A.T.M. Van Knegsel, C.E. Van Middelaar, H. Hogeveen, B. Kemp, and I.J.M. De Boer. Validation of sensor-recorded lying bouts in lactating dairy cows using a 2-sensor approach. *Journal of Dairy Science*, 98(11):7911–7916, 2015.

[19] K. Hermans, G. Opsomer, B. Van Ranst, and M. Hostens. Promises and challenges of big data associated with automated dairy cow welfare assessment. *Animal Welfare in a Changing World*, pages 199–207, 2018.

[20] C.M. Wathes, H. H. Kristensen, J-M. Aerts, and D. Berckmans. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? *Computers and electronics in agriculture*, 64(1):2–10, 2008.

[21] P. Niloofar, D. P. Francis, S. Lazarova-Molnar, A. Vulpe, M-C. Vochin, G. Suciu, M. Balanescu, V. Anestis, and T. Bartzanas. Data-driven decision support in livestock farming for improved animal health, welfare and greenhouse gas emissions: Overview and challenges. *Computers and Electronics in Agriculture*, 190:106406, 2021.

[22] T. Huybrechts, K. Mertens, J. De Baerdemaeker, B. De Ketelaere, and W. Saeys. Early warnings from automatic milk yield monitoring with online synergistic control. *Journal of Dairy Science*, 97(6):3371–3381, 2014.

[23] I. Adriaens, W. Saeys, T. Huybrechts, C. Lamberigts, L. François, K. Geerinckx, J. Leroy, B. De Ketelaere, and B. Aernouts. A novel system for on-farm fertility monitoring based on milk progesterone. *Journal of dairy science*, 101(9):8369–8382, 2018.

8 Appendices

Table 4: Statistical and non-statistical features calculated from the segments

feature name	categorical	description
inslatted	1	cow is >85% of the time in the slatted flooring area
seglength	0	length of the segment (in time)
maxgapsize	0	maximum gap size of the data in the segment
gappercent	0	percentage of the segment in time without data
nextseggap	0	gapsize of the next segment
avgdifoutlZ	0	difference between the normalised Z level of the current and the previous segment, excluding outliers
avgdifoutlCD	0	difference between the normalised CD level of the current and the previous segment, excluding outliers
rangeZ	0	range of the normalised Z values of the segment
rangeCD	0	range of the normalised CD values of the segment
difquantrangeZ	0	difference between the interquantile (5-95%) range and the full range of the normalised Z data
difquantrangeCD	0	difference between the interquantile (5-95%) range and the full range of the normalised CD data
avgoutlZ	0	average (i.e., level) of the normalised Z data without outliers
avgoutlCD	0	average (i.e., level) of the normalised CD data without outliers
stdoutlZ	0	standard deviation of normalised Z data without outliers
stdoutlCD	0	standard deviation of normalised CD data without outliers
outlpercentZ	0	percentage of outliers in the normalised Z data of the segment
outlpercentCD	0	percentage of outliers in the normalised CD data of the segment