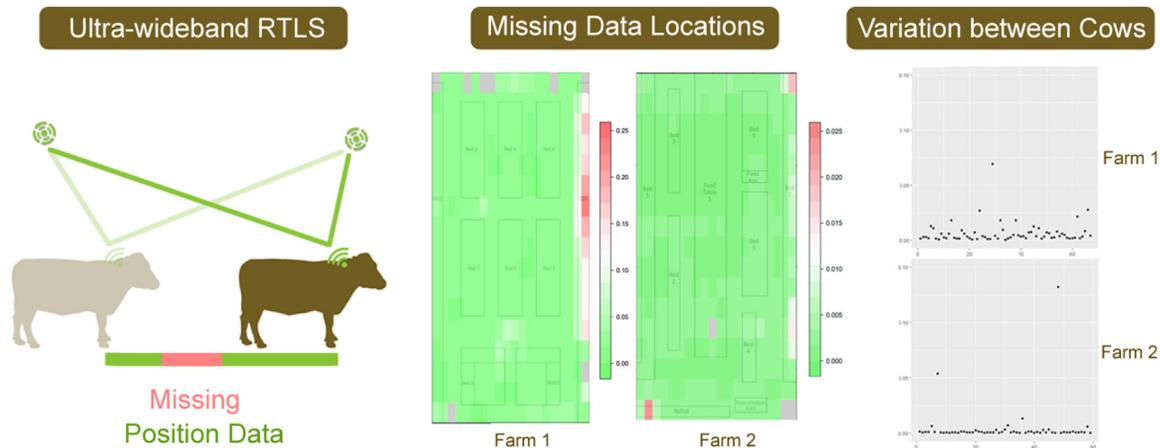


# Where do we find missing data in a commercial real-time location system? Evidence from 2 dairy farms

Keni Ren,<sup>1\*</sup> Per Peetz Nielsen,<sup>2</sup> Moudud Alam,<sup>3</sup> and Lars Rönnegård<sup>1,3</sup>

## Graphical Abstract

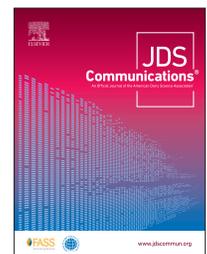


## Summary

This study analyzed the missing data in a commercial real-time location system (RTLS) within 2 dairy barns. Higher proportions of missing data were found along one of the walls on both farms. There was significant variation between individual tags.

## Highlights

- Data quality was examined in an ultra-wideband-based positioning system.
- Missing data patterns were examined on 2 dairy farms.
- Higher proportions of missing data were found along one of the walls on both farms.
- Our findings have implications for detailed analysis of cow interactions.



<sup>1</sup>Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden, <sup>2</sup>Department of Agriculture and Food, RISE Research Institutes of Sweden, 22363 Lund, Sweden, <sup>3</sup>School of Information and Engineering/Statistics, Dalarna University, 79131 Falun, Sweden.  
\*Corresponding author: [keni.ren@slu.se](mailto:keni.ren@slu.se). © 2021, The Authors. Published by Elsevier Inc. and FASS Inc. on behalf of the American Dairy Science Association®.  
This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>). Received December 16, 2020. Accepted June 07, 2021.

# Where do we find missing data in a commercial real-time location system? Evidence from 2 dairy farms

Keni Ren,<sup>1\*</sup>  Per Peetz Nielsen,<sup>2</sup>  Moudud Alam,<sup>3</sup>  and Lars Rönnegård<sup>1,3</sup> 

**Abstract:** Real-time indoor positioning using ultra-wideband devices provides an opportunity for modern dairy farms to monitor the behavior of individual cows; however, missing data from these devices hinders reliable continuous monitoring and analysis of animal movement and social behavior. The objective of this study was to examine the data quality, in terms of missing data, in one commercially available ultra-wideband-based real-time location system for dairy cows. The focus was on detecting major obstacles, or sections, inside open freestall barns that resulted in increased levels of missing data. The study was conducted on 2 dairy farms with an existing commercial real-time location system. Position data were recorded for 6 full days from 69 cows on farm 1 and from 59 cows on farm 2. These data were used in subsequent analyses to determine the locations within the dairy barns where position data were missing for individual cows. The proportions of missing data were found to be evenly distributed within the 2 barns after fitting a linear mixed model with spatial smoothing to logit-transformed proportions (mean = 18% vs. 4% missing data for farm 1 and farm 2, respectively), with the exception of larger proportions of missing data along one of the walls on both farms. On farm 1, the variation between individual tags was large (range: 9–49%) compared with farm 2 (range: 12–38%). This greater individual variation of proportions of missing data indicates a potential problem with the individual tag, such as a battery malfunction or tag placement issue. Further research is needed to guide researchers in identifying problems relating to data capture problems in real-time monitoring systems on dairy farms. This is especially important when undertaking detailed analyses of animal movement and social interactions between animals.

Precision livestock farming includes the use of technologies that enable the monitoring of individual animals. Real-time location systems (RTLS) can provide information about the position of individual cows inside dairy barns and allow their behavior and activity patterns to be recorded (Huhtala et al., 2007). Ultra-wideband (UWB) is one of the most reliable and accurate technologies available in the field of indoor positioning (Alarifi et al., 2016). Ultra-wideband technology can spread radio energy over a wide frequency band with a very low energy level and transmit large amounts of data while consuming little energy (Ghavami et al., 2007). The short pulses and time resolution make UWB signals less sensitive to reflections from the surroundings (Alarifi et al., 2016) and put less demand on battery life compared with other technologies (Svalastog, 2007). These features make UWB a suitable technology for indoor tracking of livestock (Porto et al., 2014; Tullo et al., 2016), for estimating feeding time (Pastell and Frondelius, 2018), and for detecting estrus in dairy cows (Arcidiacono et al., 2018). It can also be used to improve our understanding of social dynamics and connections in groups of animals (Rocha et al., 2020).

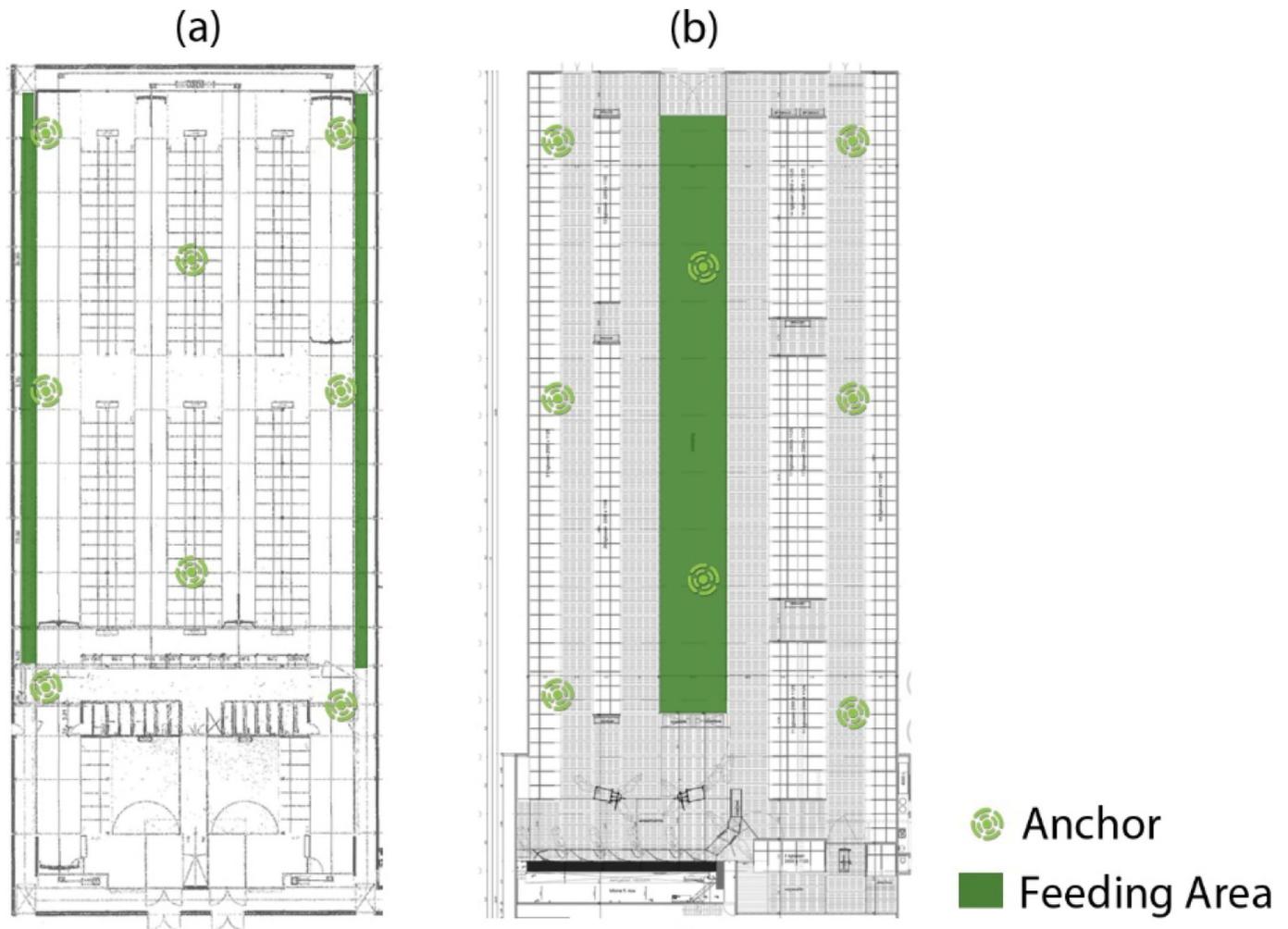
One existing commercial system based on UWB technology is the GEA CowView system (GEA Farm Technologies). The CowView system is an automatic indoor localization system for dairy cattle, providing data on positions and zone-related behavior or activity of individual animals (Sloth and Frederiksen, 2019). The CowView system was evaluated with an accuracy of 95% on detecting zone-related behavioral activities based on positioning (Tullo et al., 2016). Another UWB system with ear-tags (Ubisense) was evaluated on 8 dairy cows and obtained a mean loca-

tion accuracy within 0.52 m of the device's true location and an identification accuracy of 98% (Porto et al., 2014). Two methods have been developed to improve the accuracy of UWB positioning systems in general: by designing filters (Pastell et al., 2018) or by image analysis (Meunier et al., 2018). However, missing data, which is one of the largest problems for data preprocessing in an Internet of Things (IoT) architecture, has rarely been quantified in the livestock production research literature. In an IoT system, data are collected at high frequency from embedded sensors and shared between different resources. The low frequency of UWB pulses enables the signal to pass through walls, but metal materials can reflect part of the radio waves or block the radio waves (Liu et al., 2007). Occasional data gaps occur because of signal shaded by physical structures and calibration or transformation failures between resources (Liu et al., 2020). Moreover, missing data hinders the reliable continuous monitoring of animal activity, and missing significant amounts of data or having long time gaps will affect analysis of detailed animal movement and social behavior.

The objective of this study was to investigate whether there are any major obstacles, or sections, inside open freestall barns that would interfere with data captured from a UWB-based RTLS. More specifically, we explored the extent to which missing data in the CowView automatic monitoring system could be attributed to the cows' locations within 2 dairy barns. We also investigated variation in missing data between cows. This objective was chosen with the aim of addressing issues of interest to other researchers working on analysis of detailed animal movement and social behavior.

<sup>1</sup>Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden, <sup>2</sup>Department of Agriculture and Food, RISE Research Institutes of Sweden, 22363 Lund, Sweden, <sup>3</sup>School of Information and Engineering/Statistics, Dalarna University, 79131 Falun, Sweden.

\*Corresponding author: [keni.ren@slu.se](mailto:keni.ren@slu.se). © 2021, The Authors. Published by Elsevier Inc. and FASS Inc. on behalf of the American Dairy Science Association®. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>). Received December 16, 2020. Accepted June 07, 2021.



**Figure 1.** Floorplans of the barns equipped with the CowView systems (GEA Farm Technologies). (a) Farm 1 (74 m × 33 m) has feeding alleys on both sides of the barn toward the wall and cubicles are in the central parts of the barn. (b) Farm 2 (76 m × 30 m) has one feeding alley along the middle of the barn with cubicles on both sides.

The study was conducted on 2 commercial dairy farms with different barn structures and breeds of dairy cows. Farm 1, located in Sweden, has approximately 200 milking cows (purebred Holstein Friesian or Swedish Red and crossbreds) housed in an uninsulated freestall barn. The barn has a rectangular plan of 74 m × 33 m, with feeding alleys on both sides of the barn toward the outer walls (Figure 1a). The gable of the barn is solid concrete, and the long sides are open with a plastic curtain. The cows were kept in 2 groups such that each group had access to 1 of the 2 feeding alleys. The cubicles were fitted with rubber mattresses and sawdust as the bedding material. At the end of each bed area, there was a waist-high concrete wall. A milking parlor (2 × 12 GEA Euro class 800 with Dematron 75, GEA Farm Technologies) was located inside the barn, and cows from both groups were milked twice a day and fed a TMR ad libitum, with new feed delivered 12 times a day. Farm 2 is located in the Netherlands and has around 210 milking cows (Holstein Friesian) in an uninsulated freestall barn with a rectangular plan of 76 m × 30 m and solid walls. The feeding alley is in

the middle of the barn (Figure 1b) and cubicles contained compost made of the cows' manure as bedding material. The cows were milked at least twice a day inside the barn with 2 double automatic milking machines (Mlone 5-box, GEA Farm Technologies) and were fed roughage ad libitum and concentrate from the automatic milking system according to milk production.

Both farms had existing CowView positioning systems installed that had been actively used for approximately 10 yr for health surveillance, heat detection, and identification of cows. All cows were equipped with an active tag embedded collar, with the tag mounted on top of the neck. The position of each individual cow was provided at approximately one location per second throughout the study. The RTLS had 8 anchors (static points to define the coordinate system) installed on the ceiling at the same height throughout the barn to provide coverage of the whole barn area on both farms (Figures 1a and b, respectively). Each anchor was powered and connected to the local system through an ethernet cable. The data were automatically logged by adding an external computer

to the CowView system on each farm, where it was preprocessed through a chain of data-processing modules and then transmitted to the GEA database (Sloth and Frederiksen, 2019). It resulted in an output file containing tag name, timestamp, and the (x, y) coordinates (referred to as FA data files in the CowView system). Both farms followed their normal routine without interference from the study.

The study used data from November 2019 to January 2020 (farm 1) and from January 2020 to March 2020 (farm 2). The data were downloaded directly from GEA's server on each farm. Six days (continuous 24 h on each day) evenly distributed throughout a 3-mo period were selected. From each farm, 80 individuals were randomly chosen and only cows with data from all 6 d were included in the analysis. In total, 69 cows with 80,598,108 observations from farm 1 and 59 cows with 109,287,984 observations from farm 2 were included. Data missing 1 s or more averaged 31.29% (~7.5 h/d) and 19.97% (~4.8 h/d) of the day for farms 1 and 2, respectively. Data missing for at least 5 consecutive seconds averaged 18% (~4.3 h/d) of the day on farm 1 and 4% (~1 h/d) of the day on farm 2. Time gaps of 10 min or longer constituted 4% of the day on farm 1 and 1% of the day on farm 2. The largest period of missing data for an individual cow was 18,075 s (i.e., 5 h 2 min) on farm 1 and 27,962 s (i.e., 7 h 46 min) on farm 2.

A challenge in the analysis of missing data is separating individual cow effects from spatial effects because cows with collars giving poor signals might, purely by chance, prefer locations inside the barn that have insufficient UWB reception and thereby increase the individual variation. We addressed the problem of separating individual cow effects from spatial effects by fitting a linear mixed model to proportions of missing data with both a fixed cow effect and spatial smoothing of random grid effects. We present the linear mixed model below but first we define the response variable. More than 5 s without position data from the UWB system was considered a time gap. The threshold of 5 s was chosen because time gaps shorter than 5 s are easy to interpolate. To quantify where in the barns these time gaps occurred, a  $20 \times 17$  grid was applied (i.e., with 340 grid squares) for both farms' floorplans. The proportion of lost positions was calculated for each grid square as follows:

$$p_{ijk} = \frac{z_{ijk}}{x_{ijk}},$$

where  $x_{ijk}$  is the number of data points from cow  $i$  ( $i = 1, 2, \dots, m$ ) in grid square  $j$  ( $j = 1, 2, \dots, q$ ) on day  $k$  ( $k = 1, 2, \dots, d$ );  $z_{ijk}$  is the number of times the positions were lost for more than 5 s for cow  $i$  in square  $j$  on day  $k$ , and, consequently,  $p_{ijk}$  is the proportion of lost positions for cow  $i$  in square  $j$  at day  $k$ .

To reduce effects of varying sample size, instances where  $x_{ijk} < 40$  were deleted. Moreover, a logit-transformation was applied to improve the fit of the data to a normal distribution:

$$y_{ijk} = \log \left( \frac{p_{ijk}}{1 - p_{ijk}} \right),$$

with  $p_{ijk}$  set to  $10^{-5}$  for all  $p_{ijk} = 0$ .

The linear mixed model (equation [1]) was fitted with  $y_{ijk}$  as response, cow and day as fixed effects, and grid square as random effect ( $u$ ):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad [1]$$

where  $\mathbf{y} = \{y_{ijk}\}$  is the vector of  $n$  observations;  $\boldsymbol{\beta}$  is a vector of  $q \times 1$  fixed effects (cow and day effects) of length  $m + d - 1$ ;  $\mathbf{u}$  is a vector of spatial random effects;  $\mathbf{e}$  is a vector of residuals of length  $n$ ; and  $\mathbf{X}$  and  $\mathbf{Z}$  are design matrices of size  $n \times (m + d - 1)$  and  $n \times q$ , respectively. Furthermore,  $\mathbf{u}$  and  $\mathbf{e}$  are independently multivariate normal distributed with  $u \sim N\left(0, (\mathbf{I} - \rho\mathbf{D})^{-1}\tau\right)$  and  $e \sim N\left(0, \mathbf{I}\sigma^2\right)$ , where  $\mathbf{I}$  is the identity matrix,  $\mathbf{D}$  is the  $q \times q$  neighborhood matrix specifying which grid squares are neighbors with common edges,  $\sigma^2$  is the residual variance,  $\tau$  is the spatial variance, and  $\rho$  is the spatial dependence parameter. Hence, a conditional autoregressive (CAR) model was used and fitted using the hglm package (Alam et al., 2015) in R (R Core Team, 2020) using a convergence criterion of  $10^{-8}$ ;  $P$ -values for the cow effect (i.e., with  $H_0$ : no differences between cows) and the date effect (i.e., with  $H_0$ : no differences between dates) were computed using likelihood ratio tests. Predicted proportions ( $\hat{p}_{ijk}$ ) were calculated using the inverse-logit transform:

$$\hat{p}_{ijk} = \frac{e^{\hat{y}_{ijk}}}{1 + e^{\hat{y}_{ijk}}},$$

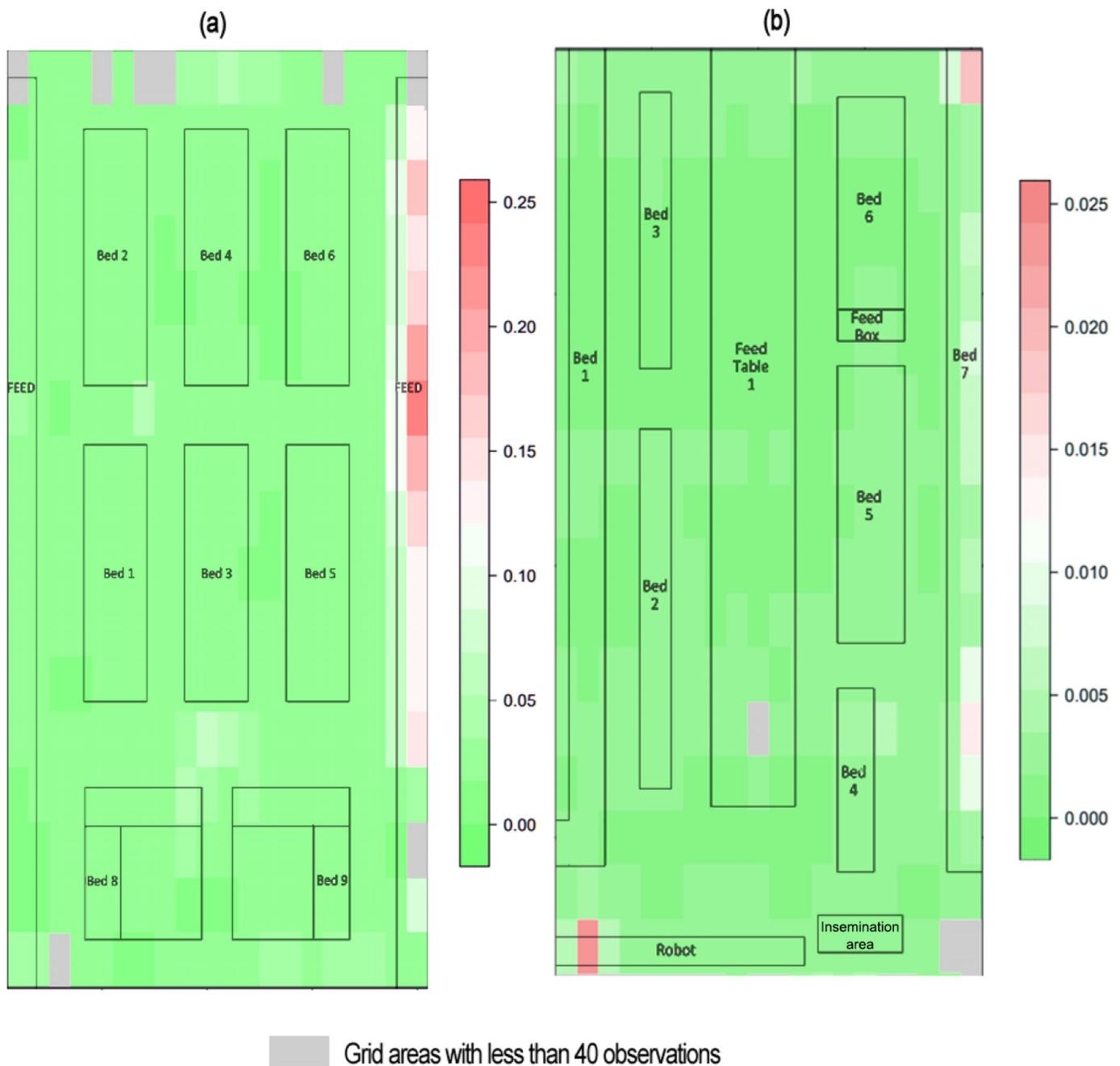
where  $\hat{y}_{ijk}$  is the fitted value from the linear mixed model (equation [1]). The maps of fitted missing proportions were computed using  $\hat{y}_{ijk}$  equal to the sum of fitted spatial random effects plus the average cow and date effects. These values are referred to as  $\hat{y}_{ijk}^*$ . The fitted missing proportions per cow were computed using  $\hat{y}_{ijk}$  equal to fitted fixed effect for that cow plus the average date effect and are referred to as  $\hat{y}_{ijk}^{**}$ .

The fitted proportions of lost positions were evenly distributed within the 2 barns (Figure 2), indicating that there were no major obstacles interfering with the RTLS inside the barns, despite them having different structures and concrete walls in different locations. However, greater amounts of data were missing along one of the walls of the barn (right-hand side) on both farms 1 and 2 (Figure 2a and 2b, respectively). This could be explained by the fact that the CowView system filters positions that are larger than the barn floorplan boundary. The levels of missing data in farm 2 were also higher at the entrance to one of the automatic milking machines (Figure 2b, red area on the left side), which was expected because of the large amounts of metal disturbing the signal and because it is in the corner of the barn, resulting in poor signal reception in this grid square.

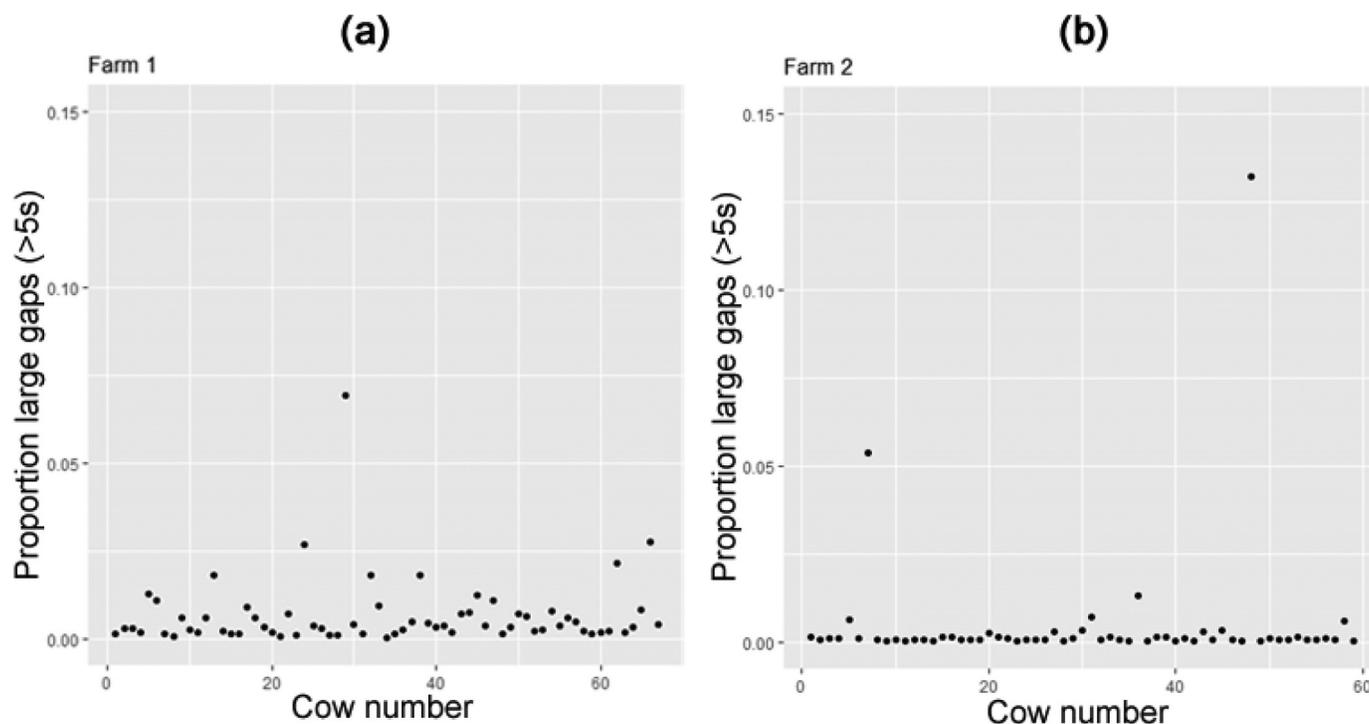
We detected large variation between the proportion of lost positions of different cows in the raw data from the CowView system (ranging between 9 and 49% in farm 1, and between 12 and 38%

in farm 2). The fitted model showed results where the variation between cows in the 2 farms was significant (both  $P < 0.01$ ), and the estimated variation between cows was larger on farm 1 than on farm 2 (Figure 3). The variation between dates was small but significant for both farms (both  $P < 0.01$ ). These individual differences could be due to battery malfunction. The low energy requirement of the tags in the collars of the CowView system ensures long

battery life (Sloth and Frederiksen, 2019); however, the uplink rate might decrease when the battery life on the collars becomes too low, resulting in more missing data. Another possible explanation for individual differences could be the position of the tag on the neck collars, resulting in a cow's neck obstructing the signal from the device. Although all collars were weighted to keep the sensor straight at the top of the neck, some sensors were observed to have



**Figure 2.** Fitted proportion of lost positions (>5 s) from the ultra-wideband system within grid squares (approximate size 4 m × 1.6 m), on (a) farm 1 and (b) farm 2. Grid squares with too few observations (gray) were excluded from the analysis. The insemination area is a closed bed area for inseminating the cows.



**Figure 3.** Estimated proportion of lost positions (>5 s) per cow from the fitted linear mixed model: (a) farm 1 and (b) farm 2. The estimated variation between cows was larger on farm 1 than on farm 2. The between-cow variation in the 2 farms was significant (both  $P < 0.01$ ).

fallen to the side of the neck or were inside out. This supports the premise that when using these devices, it is important to ensure that the tags are positioned correctly, remain in place, and are sufficiently powered at the beginning of the study.

A substantial number of time gaps along one of the walls on both farms could be problematic for accurate zone-related activity registration. Farm 1 has feeding alleys toward the 2 longer outer walls (Figure 1a), including the wall with a large number of missing positions, whereas farm 2 has feeding alleys along the middle of the barn. Therefore, feeding activity could be more difficult to follow inside the farm 1 barn. Loss of data in the feeding area is an important consideration if a researcher is interested in analyzing proximity networks and social behavior because a large number of displacements and allogrooming occur among cows in the feeding area, sometimes over short periods (Miller and Wood-Gush, 1991; Val-Laillet et al., 2009).

Filters to smooth RTLS data have been applied in earlier studies of social interactions (Chopra et al., 2020; Rocha et al., 2020), but the effect of missing data has not, to the best of our knowledge, been investigated in detail. When using a social network graph to analyze proximity interactions, missing observations of an individual can dramatically change the structure of the network graph, especially for nodes that have high betweenness (Krause et al., 2007). Because our results show significant variation between individual tags, it is important to check data quality before performing social network analysis. Furthermore, we suggest that suitable interpolation methods should be considered and validated.

In this study, we analyzed missing data from one commercially available RTLS based on UWB technology. No major obstacles

were found to interfere with the RTLS except for an automatic milking machine in one of the barns. However, large amounts of missing position data were found along one of the walls in both barns, which seems to be due to truncation of signals outside the floorplan in the CowView system rather than interference from physical objects. Our results show that although very few sections in the 2 investigated barns had high levels of missing data, the variation between individuals might have to be considered when analyzing social behavior between dairy cows, especially if these social interactions occur in an area with a high proportion of missing data.

## References

- Alam, M., L. Rönnegård, and X. Shen. 2015. Fitting conditional and simultaneous autoregressive spatial models in hglm. *R J.* 7:5–18. <https://doi.org/10.32614/RJ-2015-017>.
- Alarif, A., A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrani, M. A. Al-Ammar, and H. S. Al-Khalifa. 2016. Ultra wideband indoor positioning technologies: Analysis and recent advances. *Sensors (Basel)* 16:707. <https://doi.org/10.3390/s16050707>.
- Arcidiacono, C., S. M. Porto, M. Mancino, and G. Cascone. 2018. A software tool for the automatic and real-time analysis of cow velocity data in free-stall barns: The case study of oestrus detection from ultra-wide-band data. *Biosyst. Eng.* 173:157–165. <https://doi.org/10.1016/j.biosystemseng.2017.10.007>.
- Chopra, K., H. R. Hodges, Z. E. Barker, J. A. Vázquez Diosdado, J. R. Amory, T. C. Cameron, D. P. Croft, N. J. Bell, and E. A. Codling. 2020. Proximity interactions in a permanently housed dairy herd: Network structure, consistency, and individual differences. *Front. Vet. Sci.* 7:583715. <https://doi.org/10.3389/fvets.2020.583715>.
- Ghavami, M., L. Michael, and R. Kohno. 2007. *Ultra wideband signals and systems in communication engineering*. John Wiley & Sons.

- Huhtala, A., K. Suhonen, P. Mäkelä, M. Hakojärvi, and J. Ahokas. 2007. Evaluation of instrumentation for cow positioning and tracking indoors. *Biosyst. Eng.* 96:399–405. <https://doi.org/10.1016/j.biosystemseng.2006.11.013>.
- Krause, J., D. P. Croft, and R. James. 2007. Social network theory in the behavioural sciences: Potential applications. *Behav. Ecol. Sociobiol.* 62:15–27. <https://doi.org/10.1007/s00265-007-0445-8>.
- Liu, H., H. Darabi, P. Banerjee, and J. Liu. 2007. Survey of wireless indoor positioning techniques and systems. *IEEE Trans. Syst. Man Cybern. C* 37:1067–1080. <https://doi.org/10.1109/TSMCC.2007.905750>.
- Liu, Y., T. Dillon, W. Yu, W. Rahayu, and F. Mostafa. 2020. Missing value imputation for industrial IoT sensor data with large gaps. *IEEE Internet Things J.* 7:6855–6867. <https://doi.org/10.1109/JIOT.2020.2970467>.
- Meunier, B., P. Pradel, K. H. Sloth, C. Cirié, E. Delval, M. M. Mialon, and I. Veissier. 2018. Image analysis to refine measurements of dairy cow behaviour from a real-time location system. *Biosyst. Eng.* 173:32–44. <https://doi.org/10.1016/j.biosystemseng.2017.08.019>.
- Müller, K., and D. Wood-Gush. 1991. Some effects of housing on the social behaviour of dairy cows. *Anim. Sci.* 53:271–278. <https://doi.org/10.1017/S0003356100020262>.
- Pastell, M., and L. Frondelius. 2018. A hidden Markov model to estimate the time dairy cows spend in feeder based on indoor positioning data. *Comput. Electron. Agric.* 152:182–185. <https://doi.org/10.1016/j.compag.2018.07.005>.
- Pastell, M., L. Frondelius, M. Järvinen, and J. Backman. 2018. Filtering methods to improve the accuracy of indoor positioning data for dairy cows. *Biosyst. Eng.* 169:22–31. <https://doi.org/10.1016/j.biosystemseng.2018.01.008>.
- Porto, S. M. C., C. Arcidiacono, A. Giummarra, U. Anguzza, and G. Cascone. 2014. Localisation and identification performances of a real-time location system based on ultra wide band technology for monitoring and tracking dairy cow behaviour in a semi-open free-stall barn. *Comput. Electron. Agric.* 108:221–229. <https://doi.org/10.1016/j.compag.2014.08.001>.
- R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [www.r-project.org](http://www.r-project.org).
- Rocha, L. E., O. Terenius, I. Veissier, B. Meunier, and P. P. Nielsen. 2020. Persistence of sociality in group dynamics of dairy cattle. *Appl. Anim. Behav. Sci.* 223:104921. <https://doi.org/10.1016/j.applanim.2019.104921>.
- Sloth, K. H., and D. Frederiksen. 2019. Computer system for measuring real time position of a plurality of animals. GEA Farm Technologies GmbH, assignee. U.S. Patent No. 10,234,535.
- Svalastog, M. S. 2007. Indoor positioning-technologies, services and architectures. MS thesis. Department of Informatics, University of Oslo, Oslo, Norway. Accessed June 3, 2021. <http://urn.nb.no/URN:NBN:no-15083>.
- Tullo, E., I. Fontana, D. Gottardo, K. H. Sloth, and M. Guarino. 2016. Technical note: Validation of a commercial system for the continuous and automated monitoring of dairy cow activity. *J. Dairy Sci.* 99:7489–7494. <https://doi.org/10.3168/jds.2016-11014>.
- Val-Laillet, D., V. Guesdon, M. A. von Keyserlingk, A. M. de Passillé, and J. Rushen. 2009. Allogrooming in cattle: relationships between social preferences, feeding displacements and social dominance. *Appl. Anim. Behav. Sci.* 116:141–149. <https://doi.org/10.1016/j.applanim.2008.08.005>.

## Notes

Keni Ren  <https://orcid.org/0000-0003-2817-5331>

Per Peetz Nielsen  <https://orcid.org/0000-0002-1359-2952>

Moudud Alam  <https://orcid.org/0000-0002-3183-3756>

Lars Rønnegård  <https://orcid.org/0000-0002-1057-5401>

This project was funded by Formas—the Swedish Research Council for Sustainable Development (Stockholm, Sweden; ID: 2019-02276 and 2019-02111) and by the Kjell and Märta Beijer Foundation (Stockholm, Sweden).

The authors have not stated any conflicts of interest.