Validation!
Validation!
Validation!

2 case studies

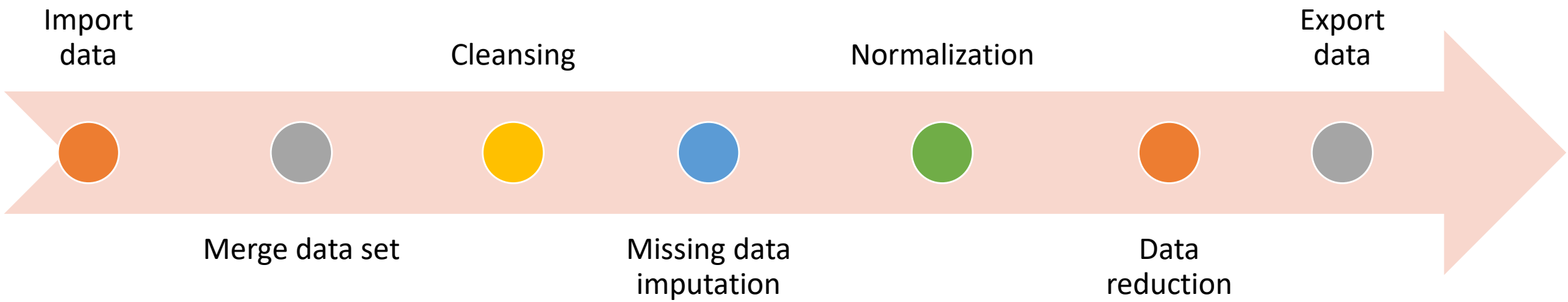
UMEÅ UNIVERSITY

Data cleaning and processing

- **Based on** Schodl, K., Stygar, A., Steininger, F., & Egger-Danner, C. (2024). Sensor data cleaning for applications in dairy herd management and breeding. *Frontiers in Animal Science*

Step	Focus
1. Validate merging	Confirm animal–sensor mapping, timestamps, system “warm-up” periods
2. Know your data	Understand units, raw/processed data, aggregation, data visualization
3. Check completeness	Identify missing or duplicated entries; consult providers to understand causes
4. Plausibility & outliers	Detect unrealistic values; reference biological plausibility
5. Technology noise	Address sensor drift, software updates, calibration issues

Preprocessing



Case 1. Field Testing Virtual Fence Collars

Why We're Testing Virtual Fence Collars

- **To verify the manufacturer's claims** – We want to see if the collars are as accurate in real-world conditions as they are on paper.
- **To protect animal welfare** – GPS errors can lead to unnecessary shocks. We test to reduce the risk of cattle being corrected when they've done nothing wrong.
- **To maximize grazing** – The more accurate the collar is near the fence, the more grass cattle can safely access—especially along the edges.

UMEÅ UNIVERSITY

What We Want to Know

How accurate the collar positions are

What happens to GNSS signals when animals move close to the virtual boundary.

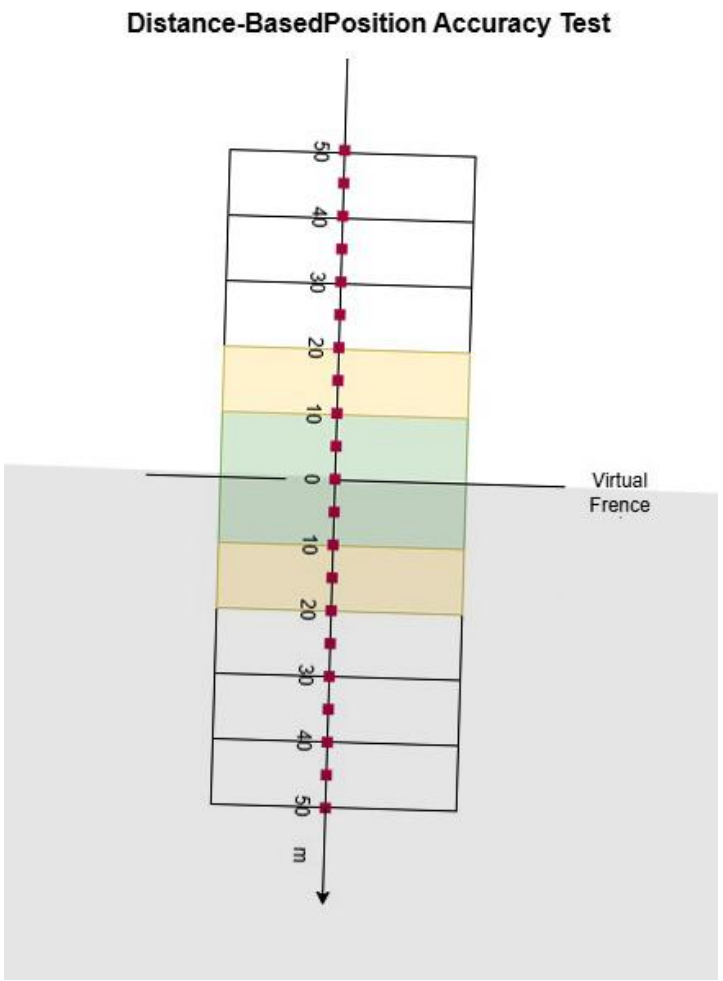
How often data is lost and whether the system reliably sends data.

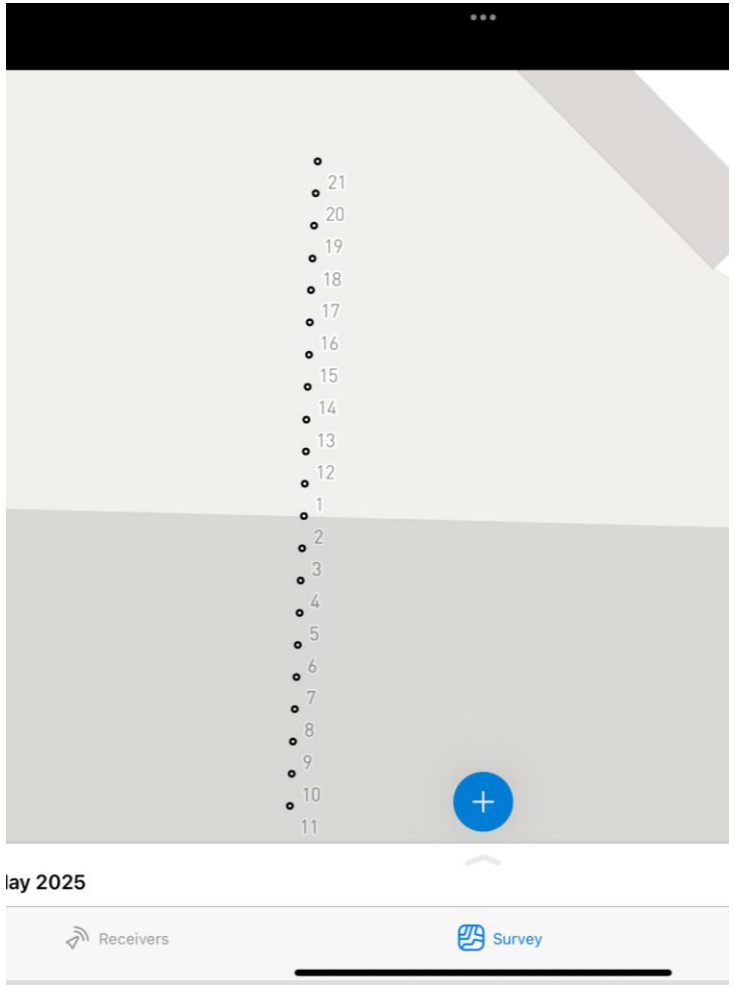
According to the Monil's technical team

- Near the boundary (within 11 meters)
  - The GPS stays **always on**, leading to **higher positioning accuracy**.
  - **Battery use is higher** due to continuous GPS activity.
- Further away from the boundary (beyond 11 meters)
  - The GPS turns **on and off** to **save battery**, which may reduce accuracy.
  - There can be a **delay when the animal moves**, as the GPS needs to "wake up."

How we do it?

- Created a 100-meter straight line — 50 meters on each side of the virtual boundary.
- Marked a point every 5 meters using high-accuracy RTK GPS.
- At each point, we paused for 10 minutes and recorded the position from the collars.

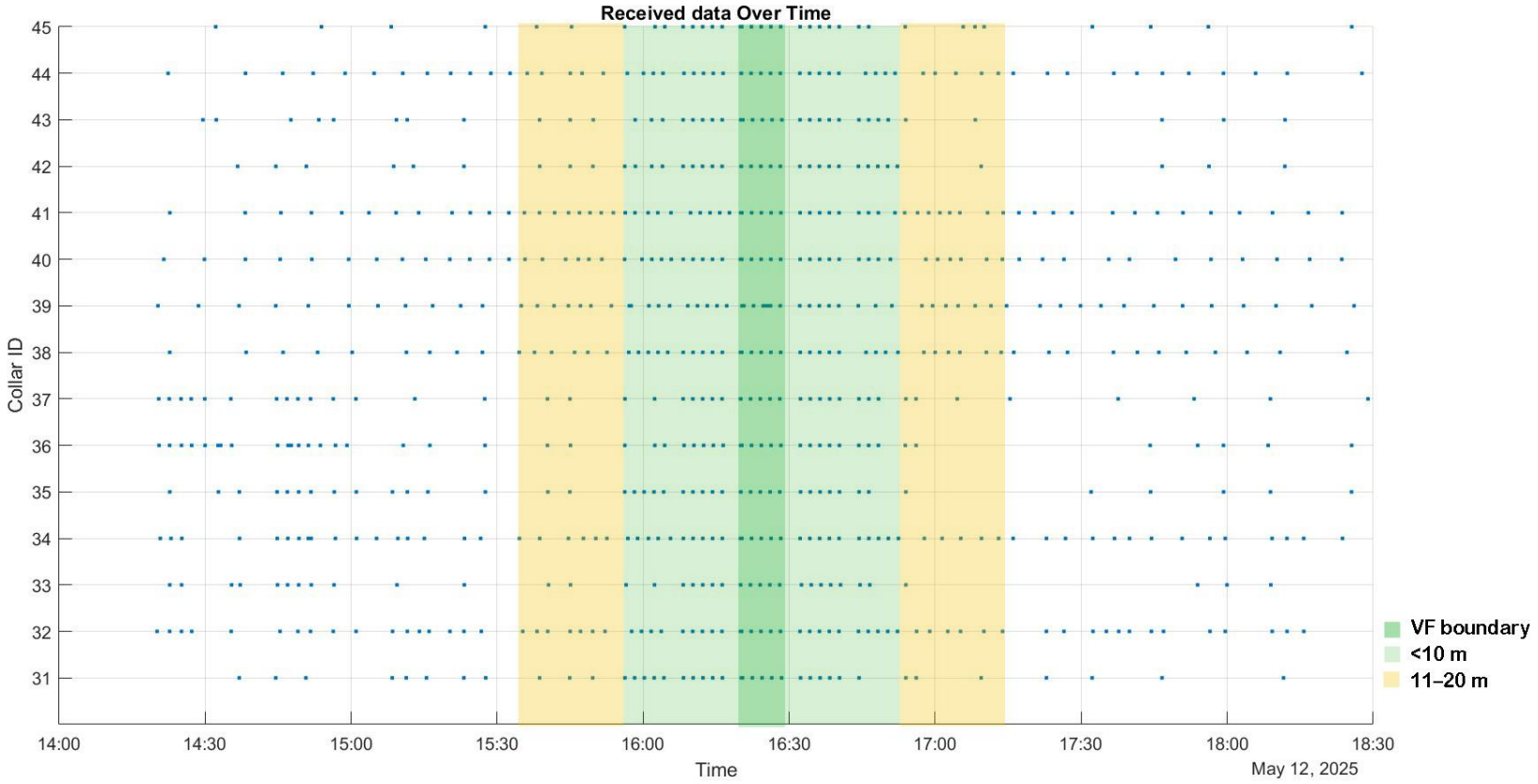




How We Collect the Data

- **15 virtual fence collars** tested across 4 farms.
- **RTK GPS unit (Emlid RS2+)** used as reference for precise positioning.
- Measured accuracy by comparing collar data to this “gold standard.”

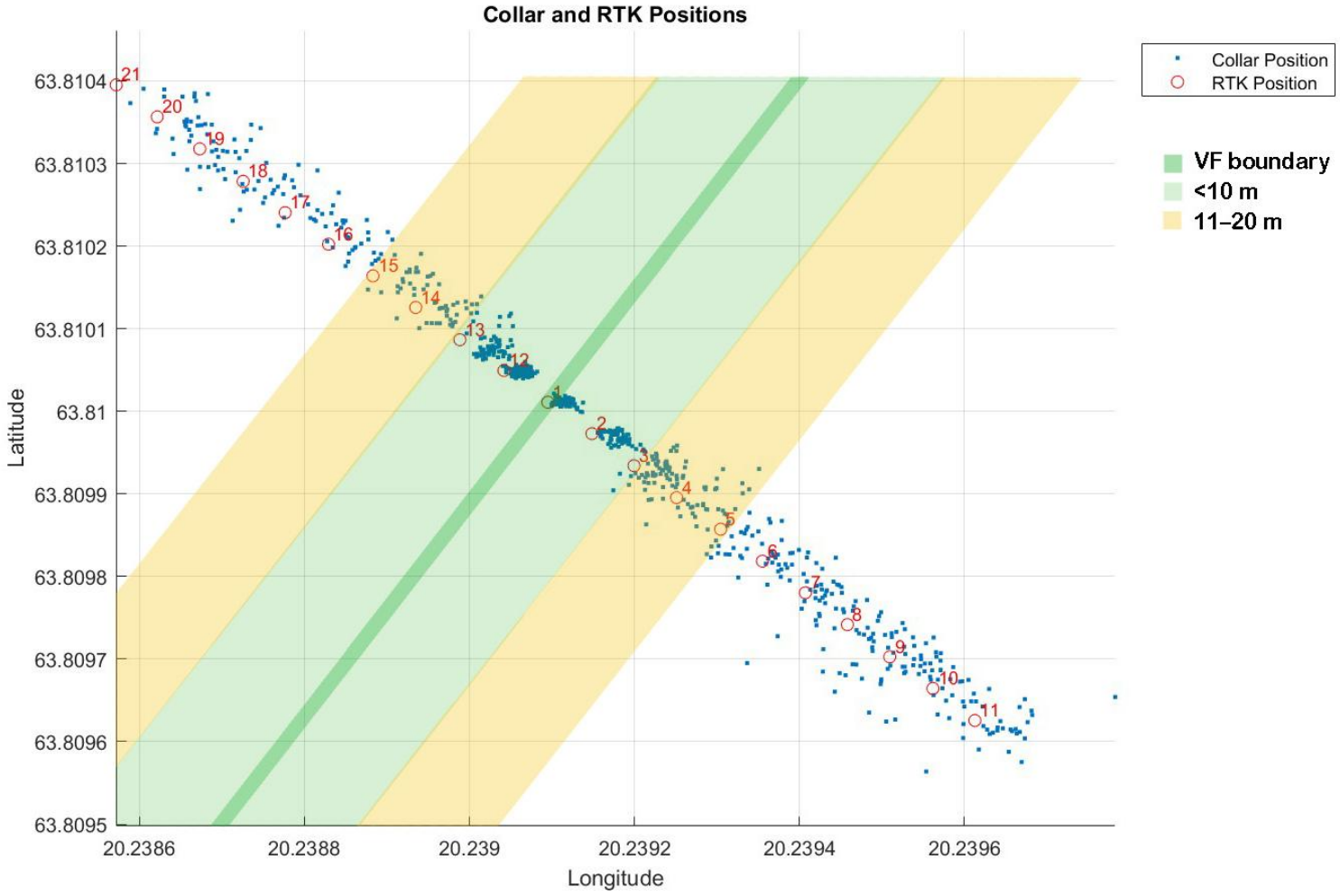
UMEÅ UNIVERSITY



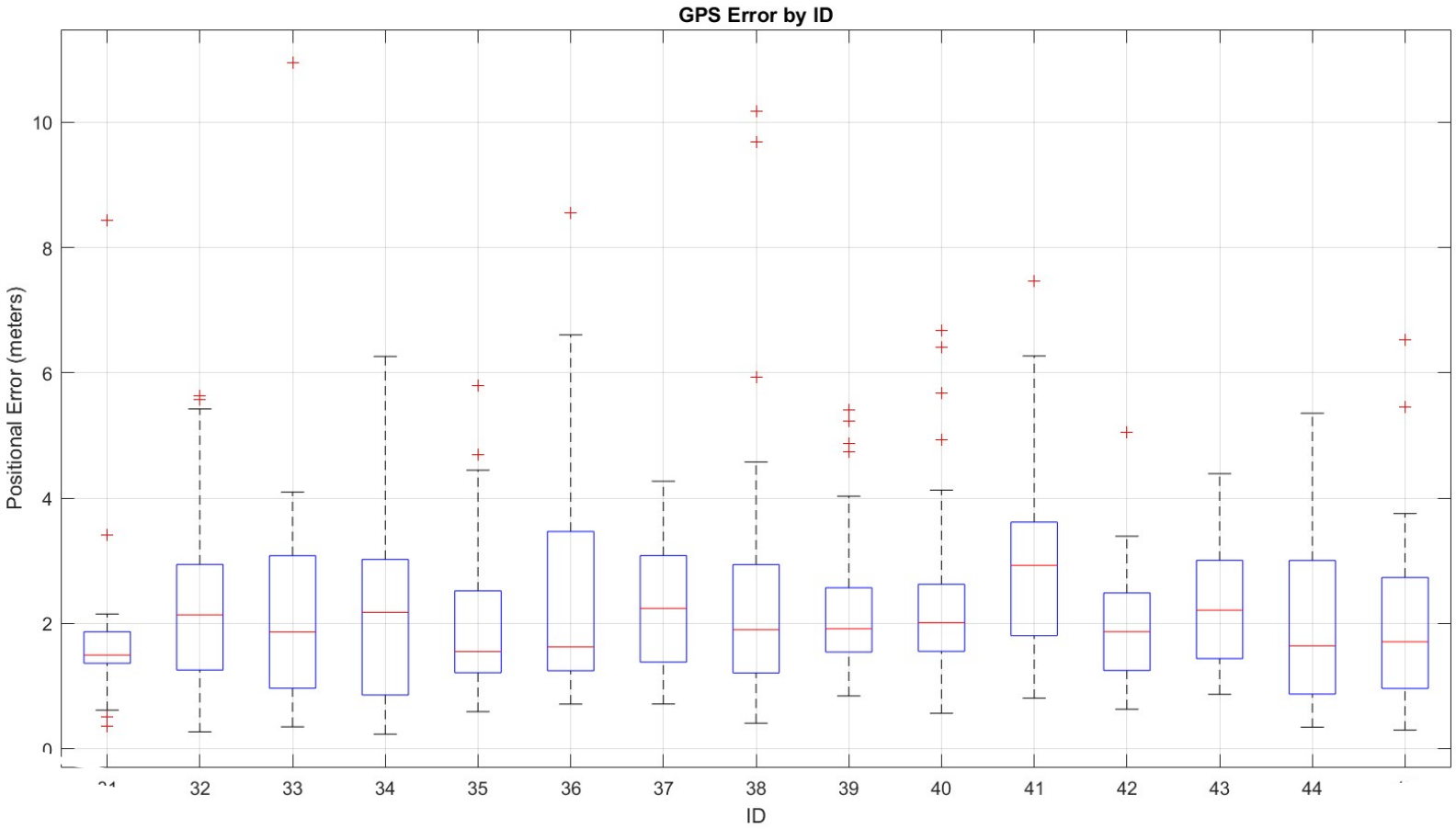
- Within 10 m, each collar registers its position every two minutes.
- Data frequency decrease as we moved further away.
- Cloud coverage and rain may affect the signal.

What Happens Near vs. Far from the Fence

UMEÅ UNIVERSITY

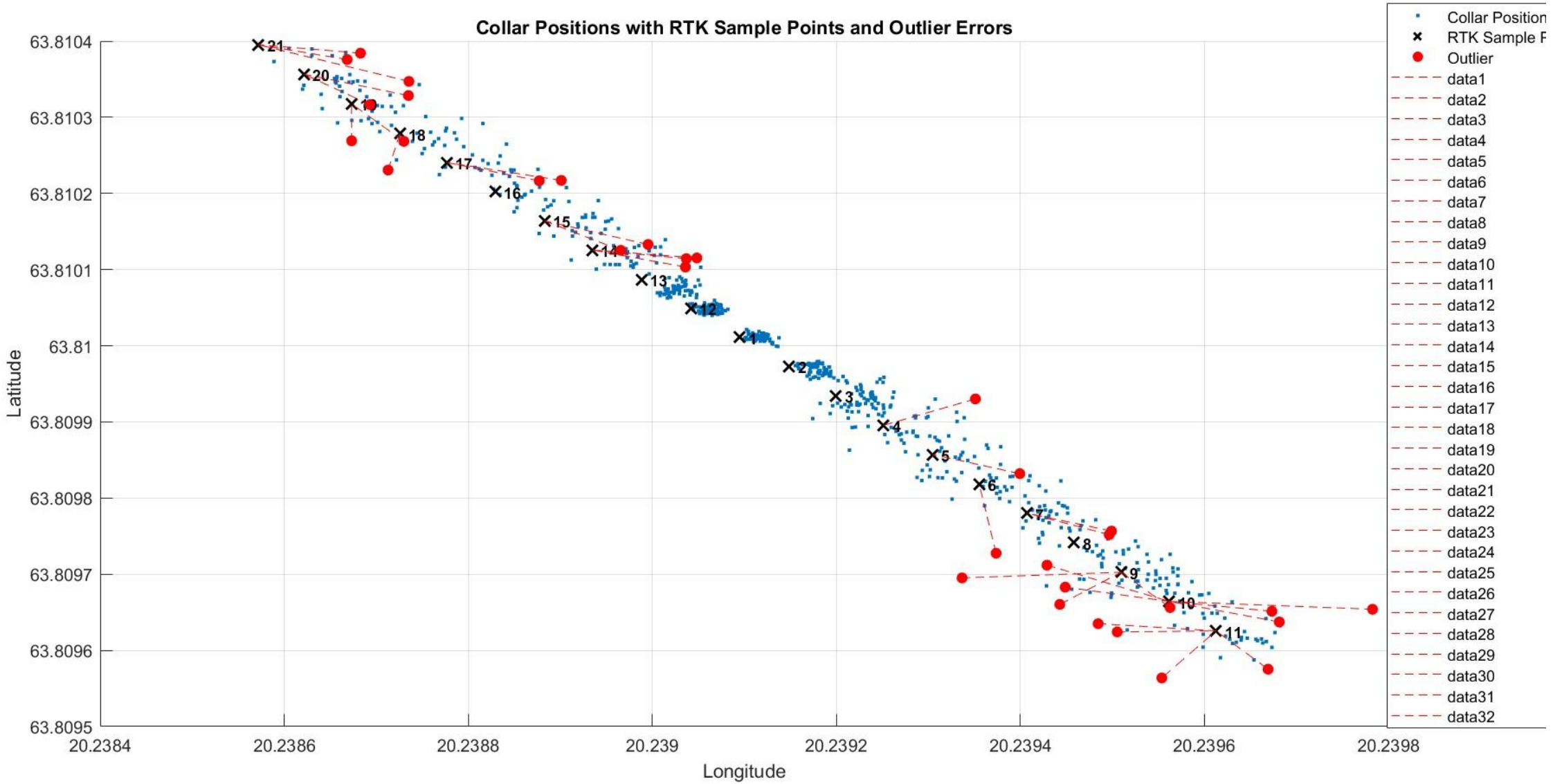


What Happens Near vs. Far from the Fence



How Accurate Were the Collars?

- Mean Error: 2.24 m
- Standard Deviation: 1.36 m


Collar Positions with RTK Sample Points and Outlier Errors
Latitude
Longitude
63.8104
63.8103
63.8102
63.8101
63.81
63.8099
63.8098
63.8097
63.8096
63.8095
20.2384
20.2386
20.2388
20.239
20.2392
20.2394
20.2396
20.2398
21
20
19
18
17
16
15
14
13
12
1
2
3
4
5
6
7
8
9
10
11
Collar Position
RTK Sample P
Outlier
data1
data2
data3
data4
data5
data6
data7
data8
data9
data10
data11
data12
data13
data14
data15
data16
data17
data18
data19
data20
data21
data22
data23
data24
data25
data26
data27
data28
data29
data30
data31
data32

What This Means

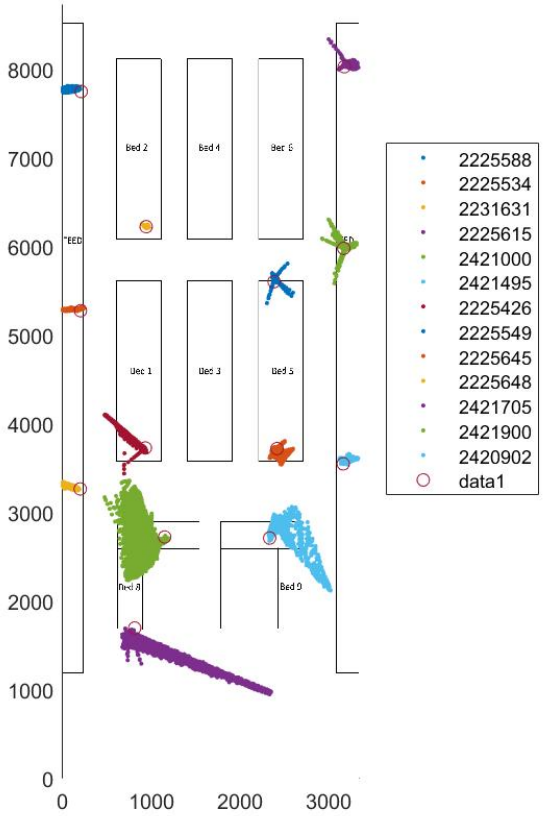
- Accuracy is generally good—but positioning gets better near the fence.
- Some data gaps can occur, especially in bad weather or poor signal areas.
- Battery lasts longer when animals are away from the fence.
- Virtual fencing can be a reliable tool, but positioning and data behaviour need to be understood for best results.

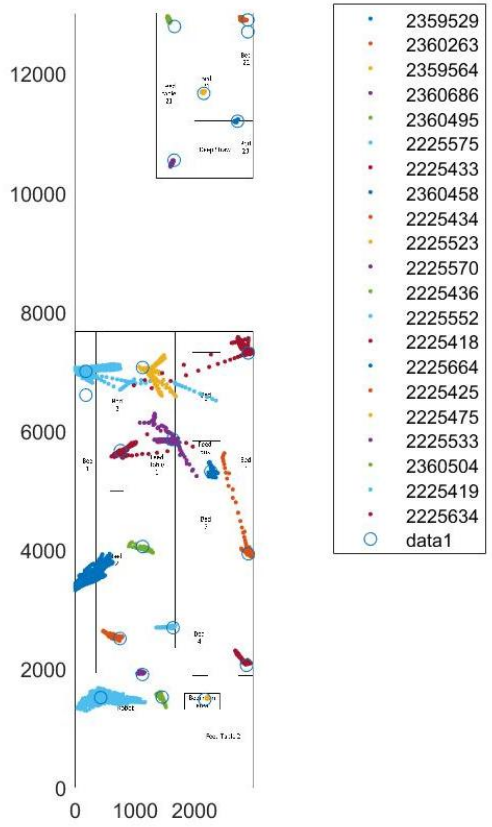
Case 2: Finding the Missing Pieces

Validation of the RTLS system

- Error distance
- Shift
- Missing data
- Activity accuracy

Error distance

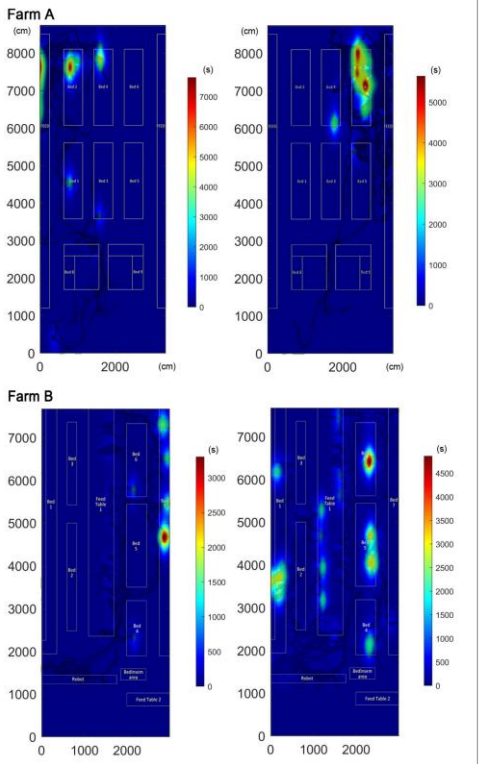


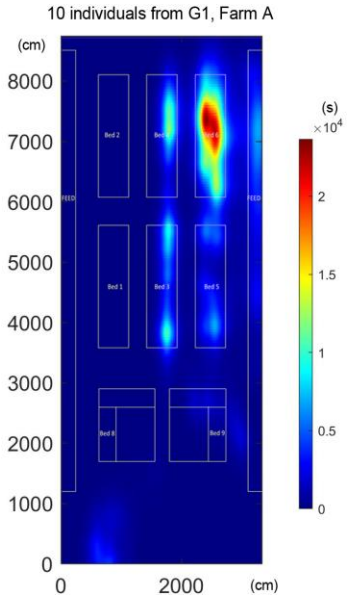


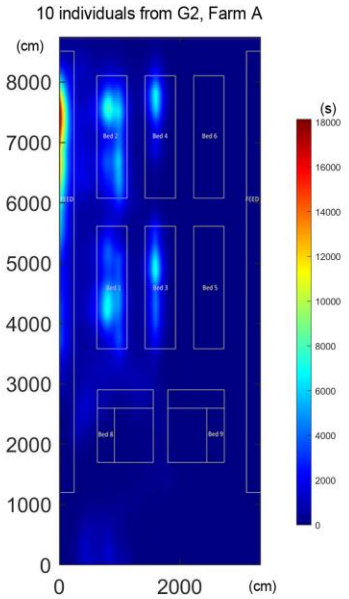
- Fixed performance tags (13 tags in farm A, 21 tags in farm B), 5 days.
- Mean error distance of 78 and 54 cm in farms A and B, SD between tags was 88 cm in farm A and 35 cm in farm B.

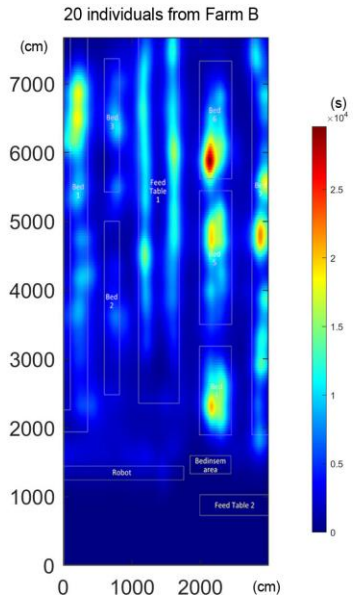
Shift of data

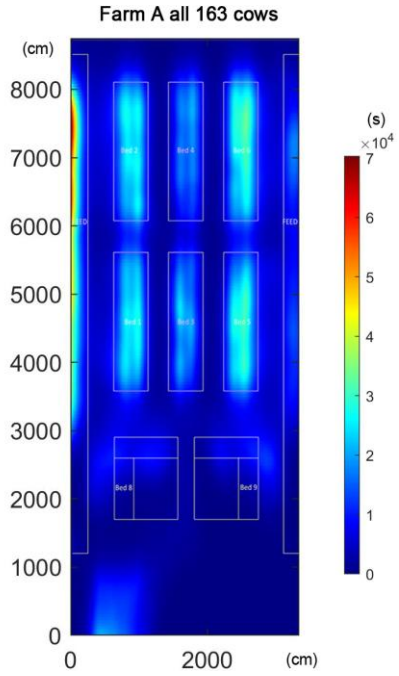
- a subset of individual
- randomly selected 20 cows from Farm A (10 from G1 and 10 cows from G2) and 20 from Farm B s
- examined the heatmap for all the cows included in our study

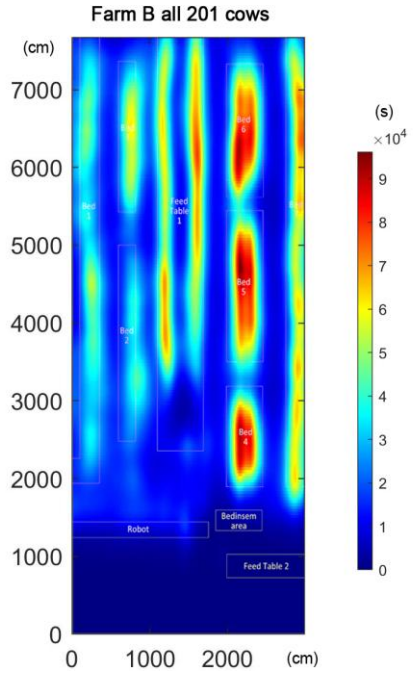












Missing Data

15-Nov-2019 00:00:15'	NaN	NaN	'15-Nov-2019 00:00:36'	NaN	NaN
'15-Nov-2019 00:00:16'	NaN	NaN	'15-Nov-2019 00:00:37'	NaN	NaN
'15-Nov-2019 00:00:17'	NaN	NaN	'15-Nov-2019 00:00:38'	3138	4102
'15-Nov-2019 00:00:18'	NaN	NaN	'15-Nov-2019 00:00:39'	NaN	NaN
'15-Nov-2019 00:00:19'	NaN	NaN	'15-Nov-2019 00:00:40'	NaN	NaN
'15-Nov-2019 00:00:20'	NaN	NaN	'15-Nov-2019 00:00:41'	3169	4122
'15-Nov-2019 00:00:21'	NaN	NaN	'15-Nov-2019 00:00:42'	NaN	NaN
'15-Nov-2019 00:00:22'	NaN	NaN	'15-Nov-2019 00:00:43'	NaN	NaN
'15-Nov-2019 00:00:23'	NaN	NaN	'15-Nov-2019 00:00:44'	NaN	NaN
'15-Nov-2019 00:00:24'	NaN	NaN	'15-Nov-2019 00:00:45'	NaN	NaN
'15-Nov-2019 00:00:25'	NaN	NaN	'15-Nov-2019 00:00:46'	NaN	NaN
'15-Nov-2019 00:00:26'	NaN	NaN	'15-Nov-2019 00:00:47'	NaN	NaN
'15-Nov-2019 00:00:27'	NaN	NaN	'15-Nov-2019 00:00:48'	NaN	NaN
'15-Nov-2019 00:00:28'	NaN	NaN	'15-Nov-2019 00:00:49'	NaN	NaN
'15-Nov-2019 00:00:29'	NaN	NaN	'15-Nov-2019 00:00:50'	NaN	NaN
'15-Nov-2019 00:00:30'	NaN	NaN	'15-Nov-2019 00:00:51'	NaN	NaN
'15-Nov-2019 00:00:31'	NaN	NaN	'15-Nov-2019 00:00:52'	NaN	NaN
'15-Nov-2019 00:00:32'	NaN	NaN	'15-Nov-2019 00:00:53'	NaN	NaN
'15-Nov-2019 00:00:33'	NaN	NaN	'15-Nov-2019 00:00:54'	NaN	NaN
'15-Nov-2019 00:00:34'	NaN	NaN	'15-Nov-2019 00:00:55'	NaN	NaN
'15-Nov-2019 00:00:35'	NaN	NaN	'15-Nov-2019 00:00:56'	NaN	NaN

1

Examining the data quality

Used the random 69 cows to determine the location and duration of missing data

K Ren, PP Nielsen, M Alam, L Rönnegård - JDS Communications, 2021

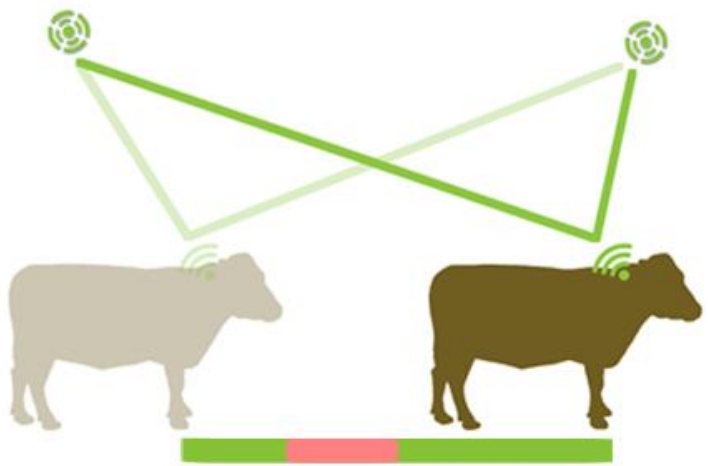
2

Work on the interpolation method.

Using the 20 cow with high performance tags to compare the effects of four different interpolation methods

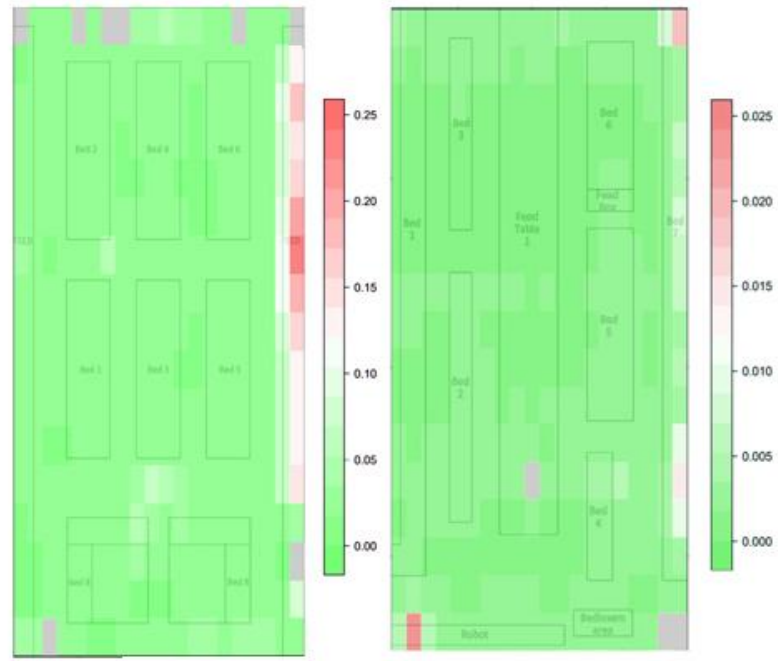
Ren, K., Alam, M., Peetz Nielsen, P., Gussmann, M., & Rönnegård, L. - *Frontiers in Animal Science,* 2022

Ultra-wideband RTLS



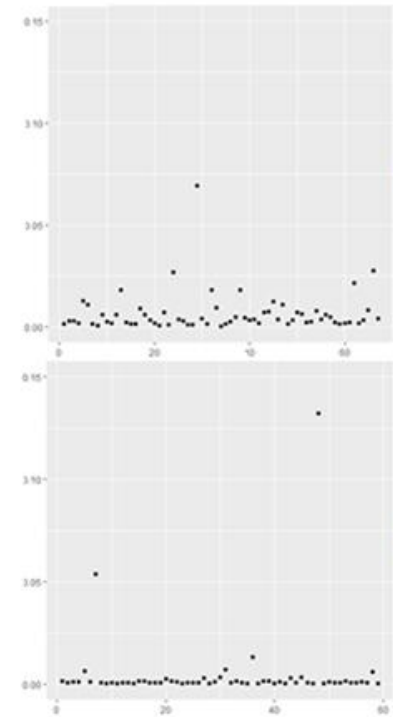
Missing
Position Data

Missing Data Locations



Farm 1 Farm 2

Variation between Cows

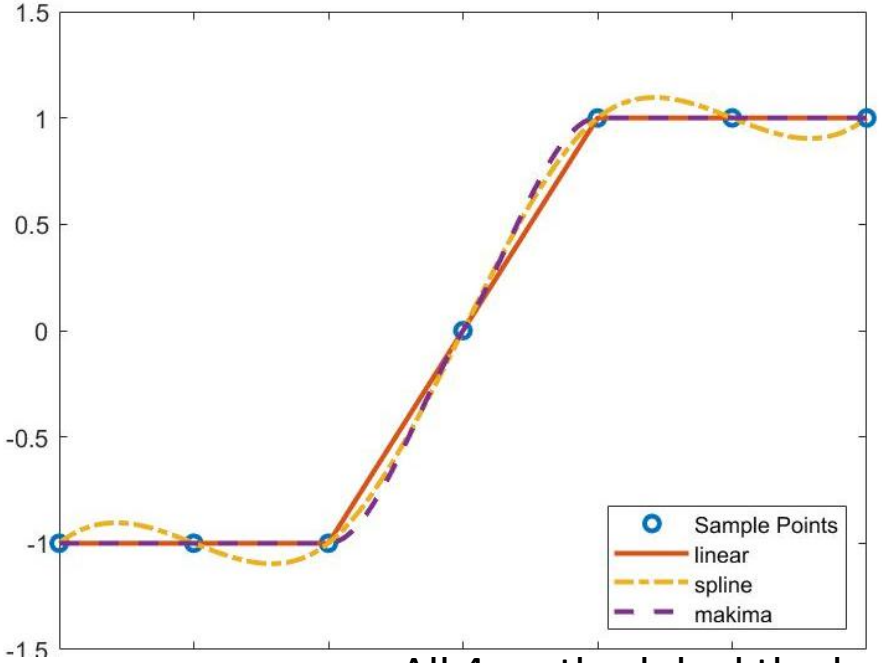


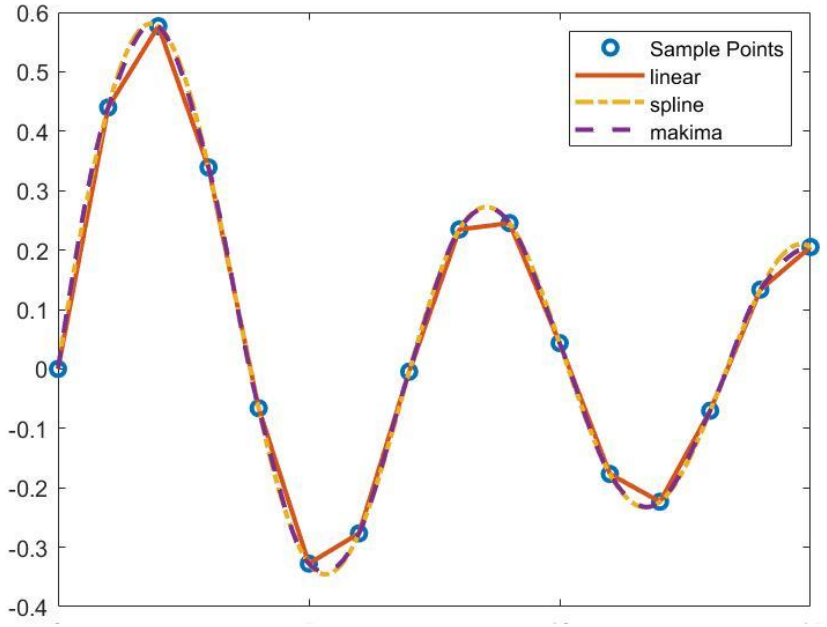
Farm 1

Farm 2

Data interpolation

- Previous
- Linear
- Spline
- Makima

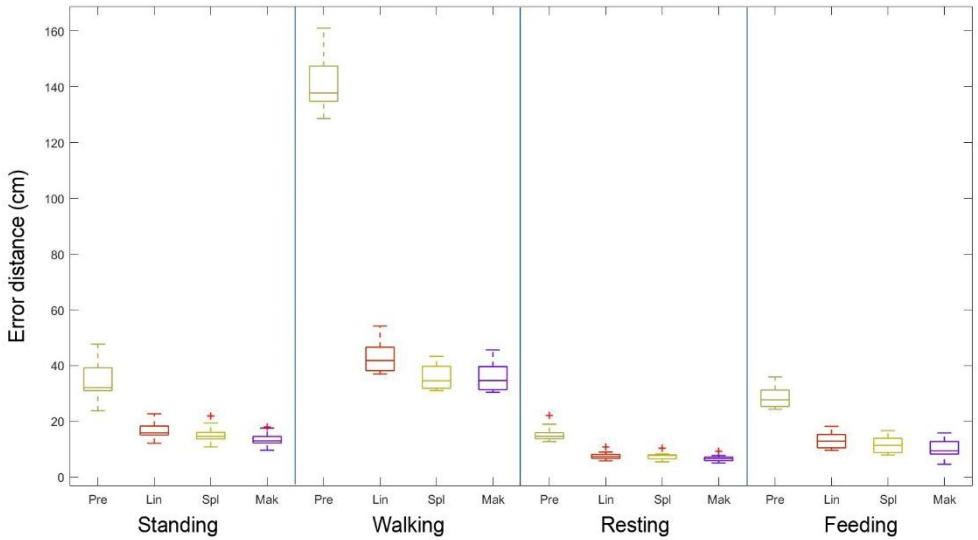


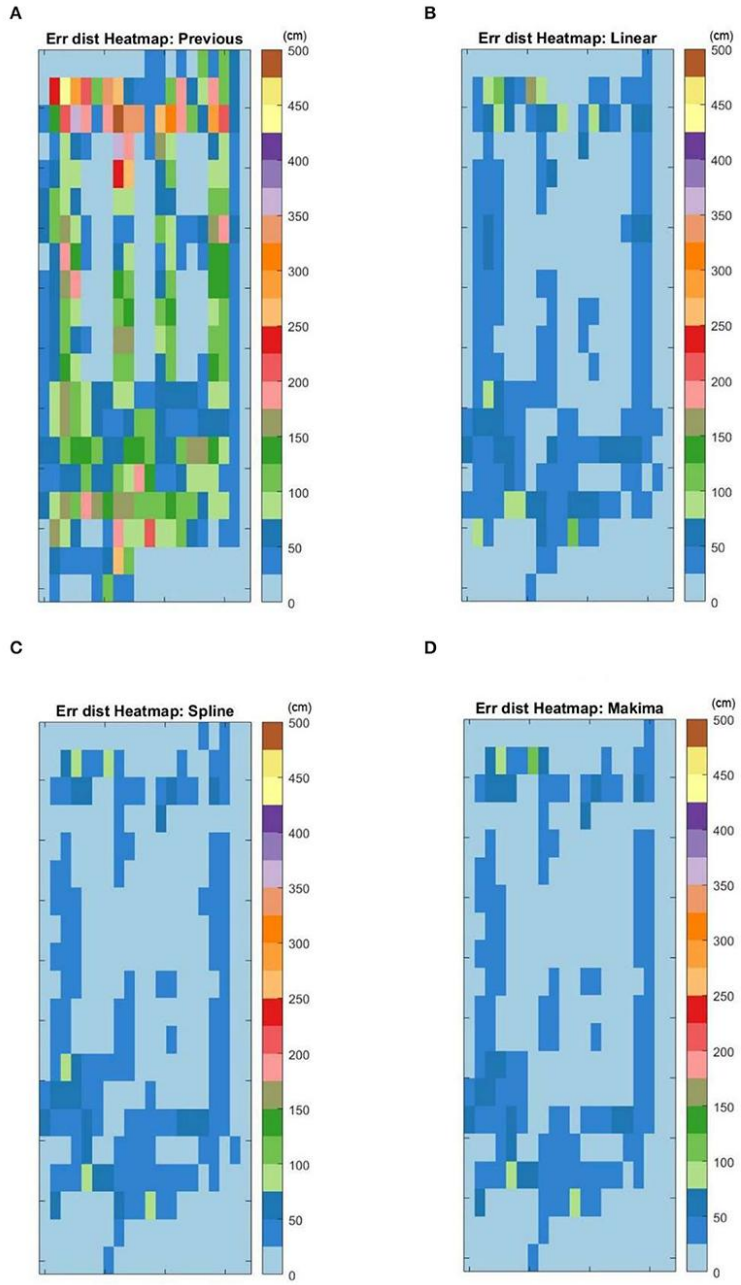


All 4 method deal the beginning missing using the first non-missing values, and end values as the previous non-missing values

All 4 method drag the interpolated data back from the barn boarder.

	Mean (cm)	Standard deviation (cm)
Previous	55.1	51.1
Linear	20.2	14.2
Spline	17.7	11.5
Makima	16.6	11.9

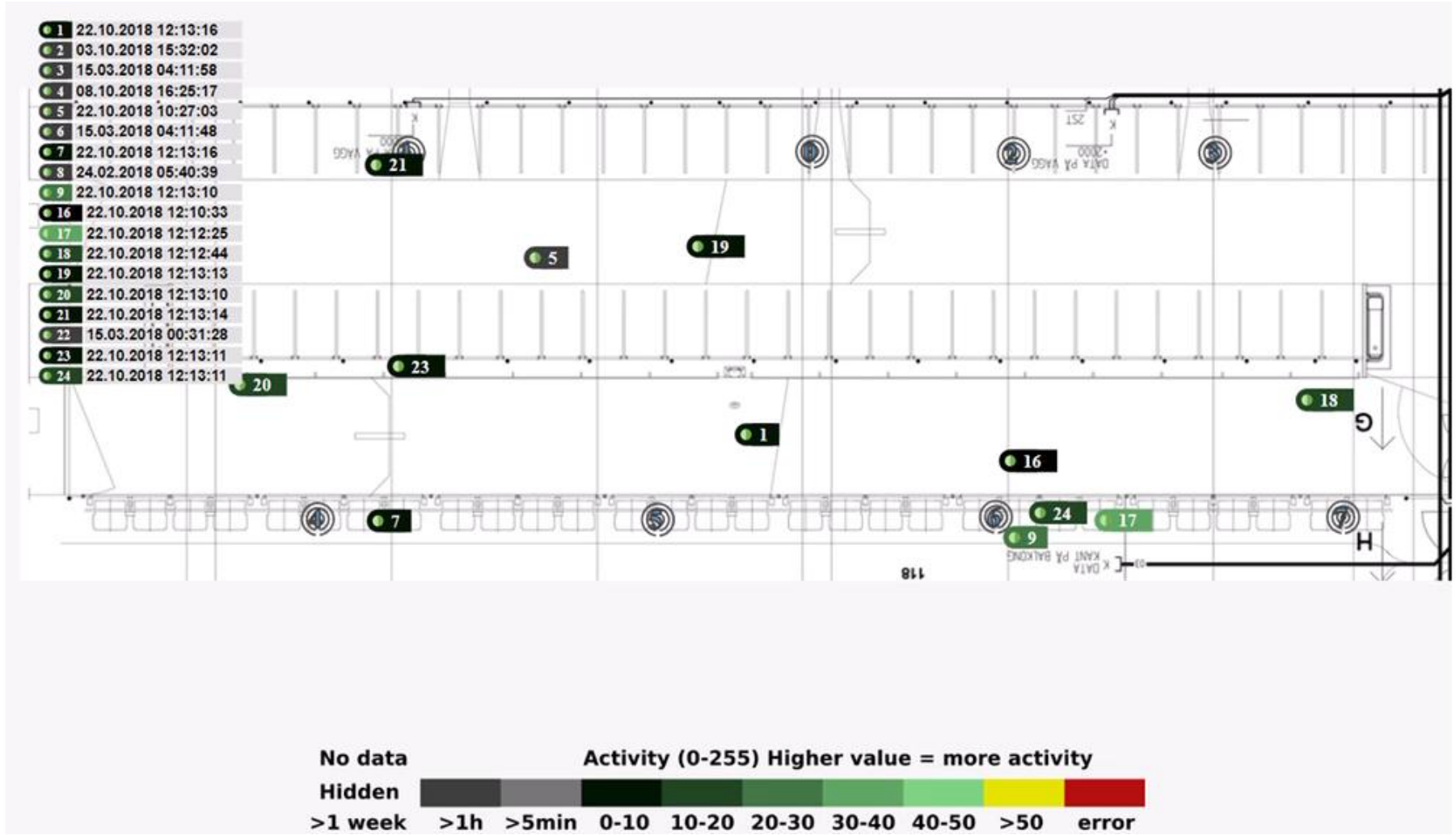




Ren, K., Alam, M., Peetz Nielsen, P., Gussmann, M., & Rönnegård, L. - *Frontiers in Animal Science,* 2022

Time Budget Analysis

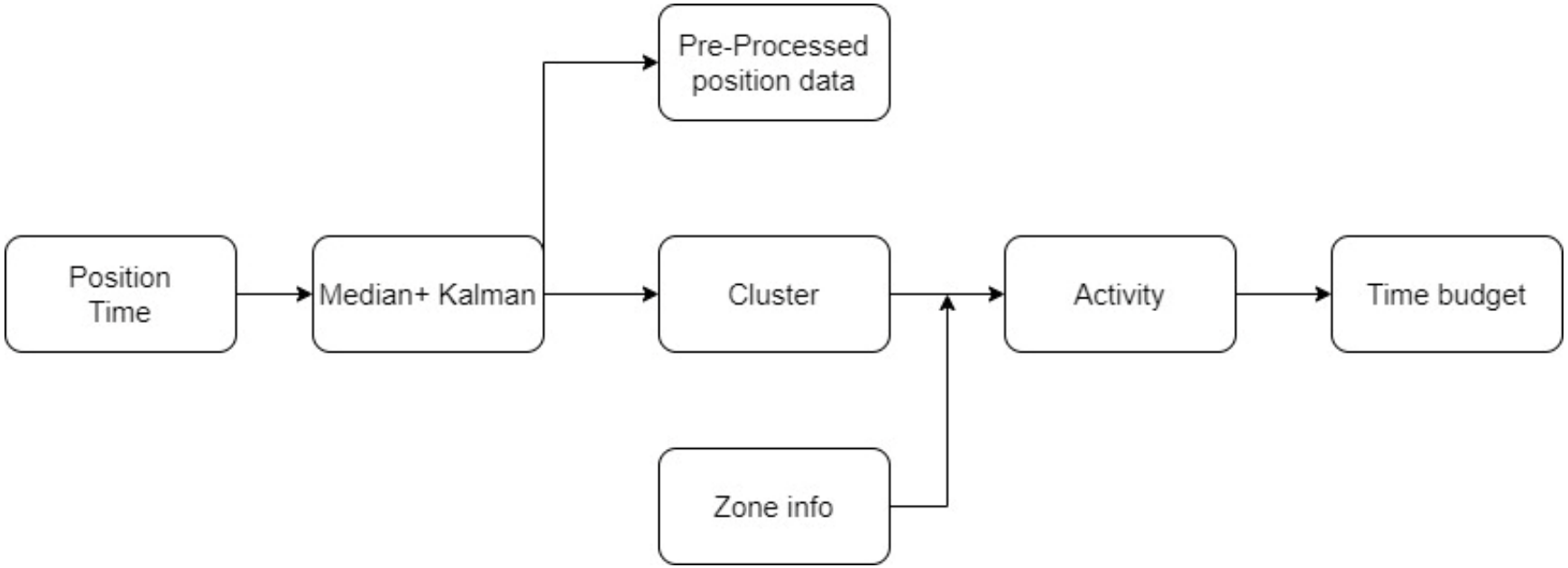
Keni Ren


1 22.10.2018 12:13:16
2 03.10.2018 15:32:02
3 15.03.2018 04:11:58
4 08.10.2018 16:25:17
5 22.10.2018 10:27:03
6 15.03.2018 04:11:48
7 22.10.2018 12:13:16
8 24.02.2018 05:40:39
9 22.10.2018 12:13:10
16 22.10.2018 12:10:33
17 22.10.2018 12:12:25
18 22.10.2018 12:12:44
19 22.10.2018 12:13:13
20 22.10.2018 12:13:10
21 22.10.2018 12:13:14
22 15.03.2018 00:31:28
23 22.10.2018 12:13:11
24 22.10.2018 12:13:11
No data
Hidden
>1 week
Activity (0-255) Higher value = more activity
>1h
>5min
0-10
10-20
20-30
30-40
40-50
>50
error

Ren, Keni, et al. "Tracking and analysing social interactions in dairy cattle with real-time locating system and machine learning." *Journal of Systems Architecture*

Indoor RTLS data

- ID, Timestamp, x, y ,(z).
- Blueprint with function zones



Data

DATA FROM POSITION SYSTEM

- FA, PA, PAA, PC
- .csv
- Everyday
- Around 800Mb/day

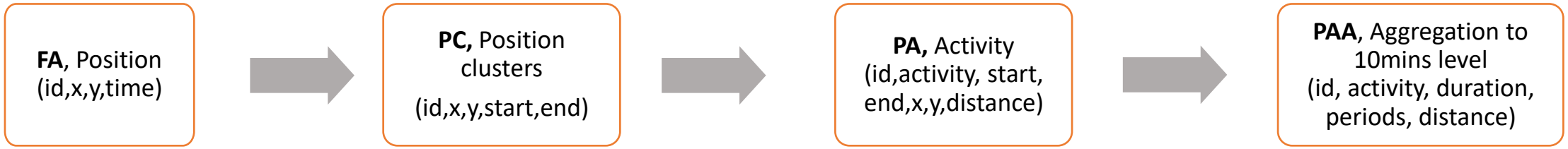
DATA FROM FARM

- Cow traits, insemination records , productions traits
- KO info, Översikt hälsotillstånd X, Avkastn 14 dag…
- .txt
- Every week or longer period

1	data_entity	tag_id	tag_string	time	x	y	z
2	FA	2428773	00250F65	1.57E+12	107	4756	198
3	FA	2428044	00250C8C	1.57E+12	13	4752	198
4	FA	2433145	252079	1.57E+12	3105	1732	198
5	FA	2428747	00250F4B	1.57E+12	2240	3098	198

KO	RESP	TAG	GR	STAT	LAKT	KALVN	DIM
601	13418550	00250cdd	11	DRÄKT	3	4-12-19	285
607	13419990	00250c8c	11	DRÄKT	3	7-01-20	251
611	14966783	00250de6	11	DRÄKT	3	15-07-19	427
659	13418482	00250dcc	21	TIDIG	3	3-09-20	11

DATA FROM POSITION SYSTEM



Sloth, Karen Helle, and Daniel Frederiksen. "Computer system for measuring real time position of a plurality of animals." U.S. Patent No. 10,234,535. 19 Mar. 2019.

FA
(id,time,x,y,z)

data_entity,tag_id,tag_string,time,x,y,z

FA,2198225,00218AD1,1427847365285,2170,2376,137

FA,2200003,002191C3,1427847361618,2827,6220,137

FA,2199224,00218EB8,1427847365333,433,6159,137

FA,2200407,00219357,1427847365379,3091,6972,137

FA,2201318,002196E6,1427847363445,404,5903,137

FA,2199144,00218E68,1427847361227,2731,2114,137

FA,2199938,00219182,1427847350225,2186,5239,137

FA,2201146,0021963A,1427847365635,1148,7545,137

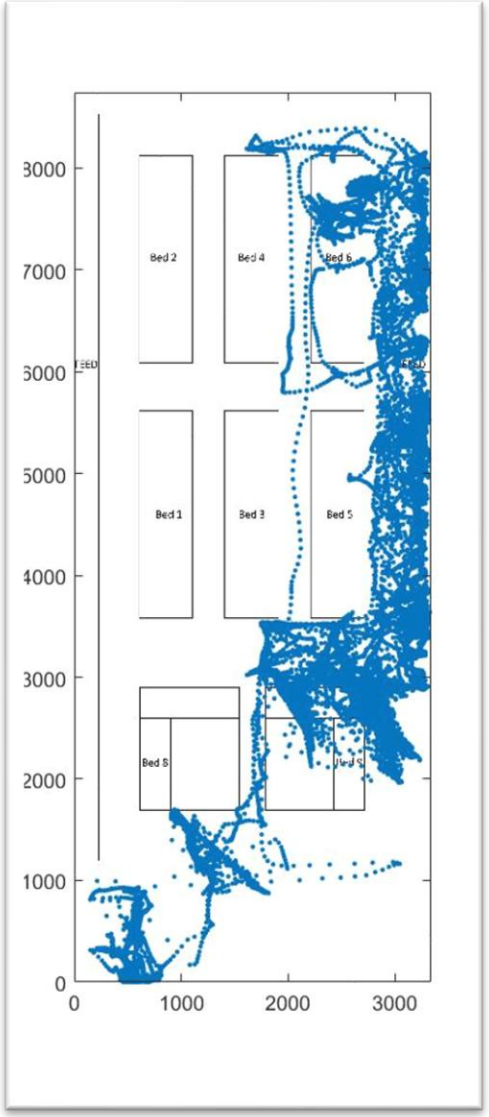
FA,2201418,0021974A,1427847355830,2799,2424,137

FA,2199316,00218F14,1427847365300,1588,3090,137

FA,2196981,002185F5,1427847360796,731,2757,137

FA,2199374,00218F4E,1427847364776,1337,3720,137

Time in epoch

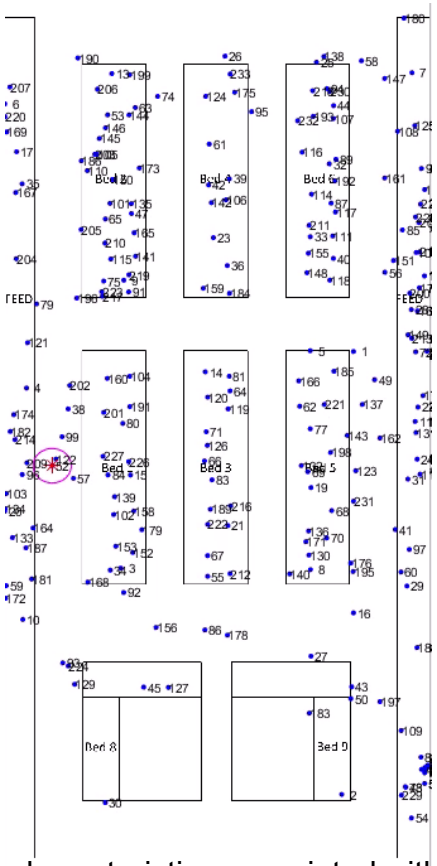


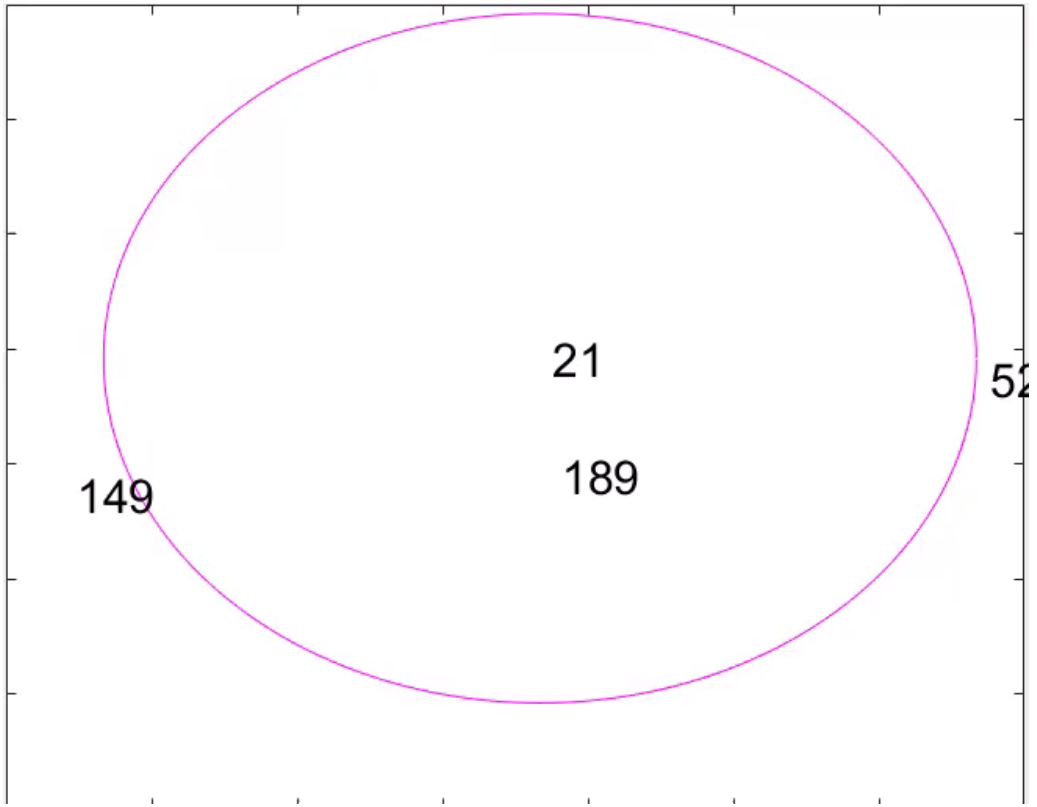
ID=2421875

15-Nov-2019 00:00:00-23:30:00

How we use FA data?

An example: Social contact





Hansson, Ida, et al (2022). Cow characteristics associated with the variation in number of contacts between dairy cows. Journal of Dairy Science.
Marina, Hector, et al. (2023) Social Network Analysis of Dairy Cow Interactions Journal of Dairy Science

PC

(id,start,end,x,y,z)

data_entitiy,tag_id,tag_string,start,end,x,y,z

PC,2198225,00218AD1,1427847365285,1427847365285,2170,2376,137

PC,2200003,002191C3,1427847361618,1427847365259,2827,6220,137

PC,2199224,00218EB8,1427847365333,1427847365333,433,6159,137

PC,2200407,00219357,1427847365379,1427847365379,3091,6972,137

PC,2201318,002196E6,1427847363445,1427847364673,404,5903,137

PC,2199144,00218E68,1427847361227,1427847364819,2731,2114,137

PC,2199938,00219182,1427847350225,1427847365511,2186,5239,137

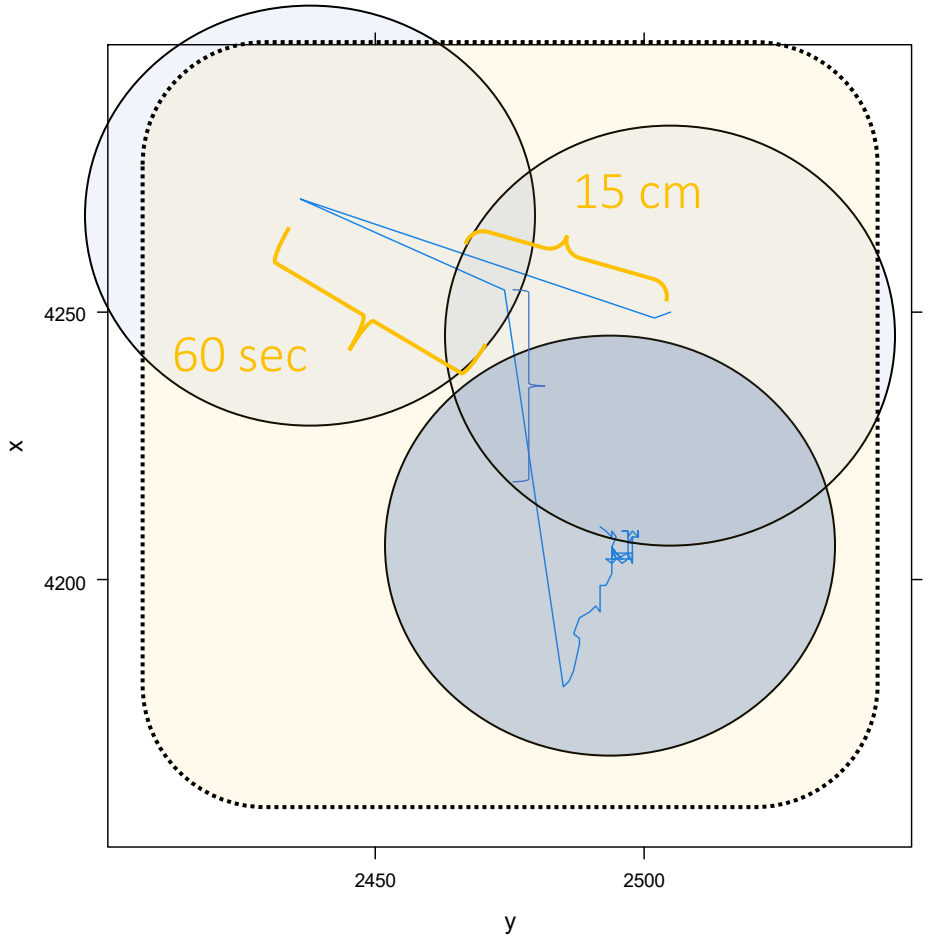
PC,2201146,0021963A,1427847365635,1427847365635,1148,7545,137

PC,2201418,0021974A,1427847355830,1427847365494,2799,2424,137

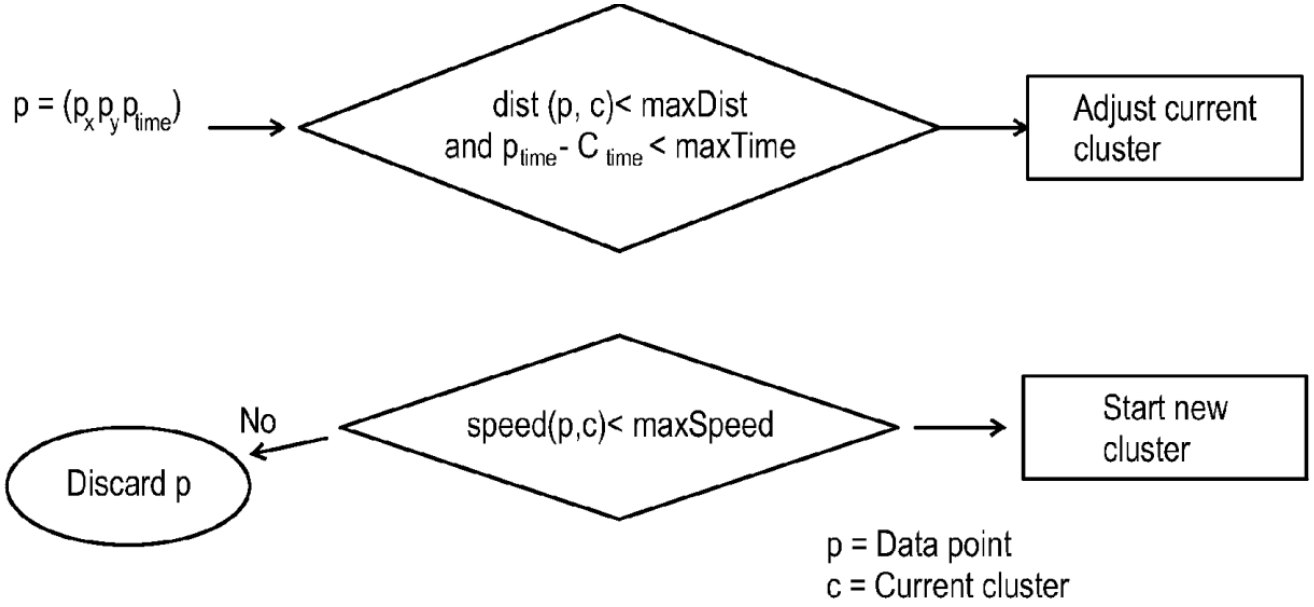
PC,2199316,00218F14,1427847365300,1427847365300,1588,3090,137

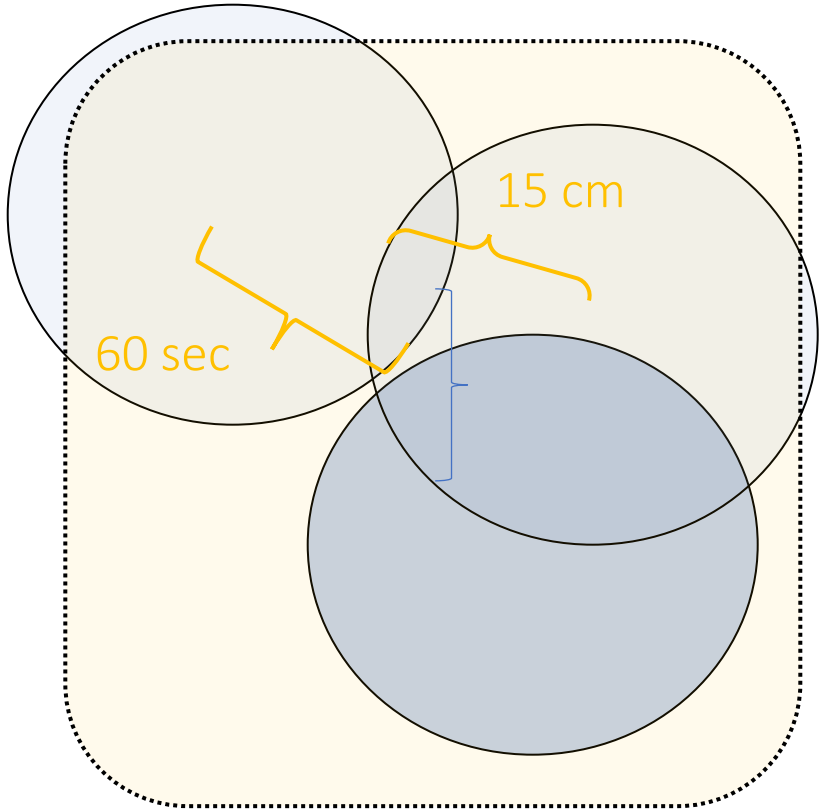
PC,2196981,002185F5,1427847360796,1427847360796,731,2757,137

PC,2199374,00218F4E,1427847364776,1427847365923,1337,3720,137



New cluster if radius > 15 cm or time between two points > 60 sec.





New cluster if radius > 15 cm or time between two points > 60 sec.

PA
(id, start, end, x, y, z, activity type, distance)

data_entity,tag_id,tag_string,start,end,x,y,z,activity_type,distance

PA,2196869,00218585,1427847353100,1427847376698,3222,5504,137,1,0

PA,2196869,00218585,1427847376698,1427847379119,3221,5522,137,2,25

PA,2196869,00218585,1427847379119,1427847381544,3221,5522,137,1,0

PA,2196869,00218585,1427847381544,1427847383959,3202,5490,137,2,15

PA,2196869,00218585,1427847383959,1427847385171,3202,5490,137,1,0

PA,2196869,00218585,1427847385171,1427847393644,3258,5679,137,2,52

PA,2196869,00218585,1427847393644,1427847413007,3277,5702,137,1,0

PA,2196869,00218585,1427847413007,1427847416637,3230,5686,137,2,26

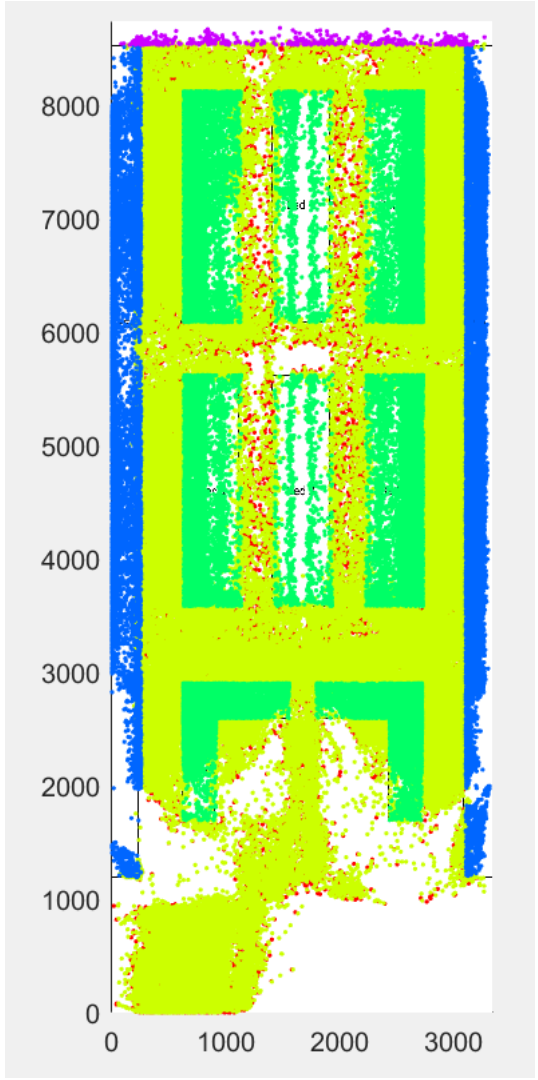
PA,2196869,00218585,1427847416637,1427847437212,3206,5662,137,1,0

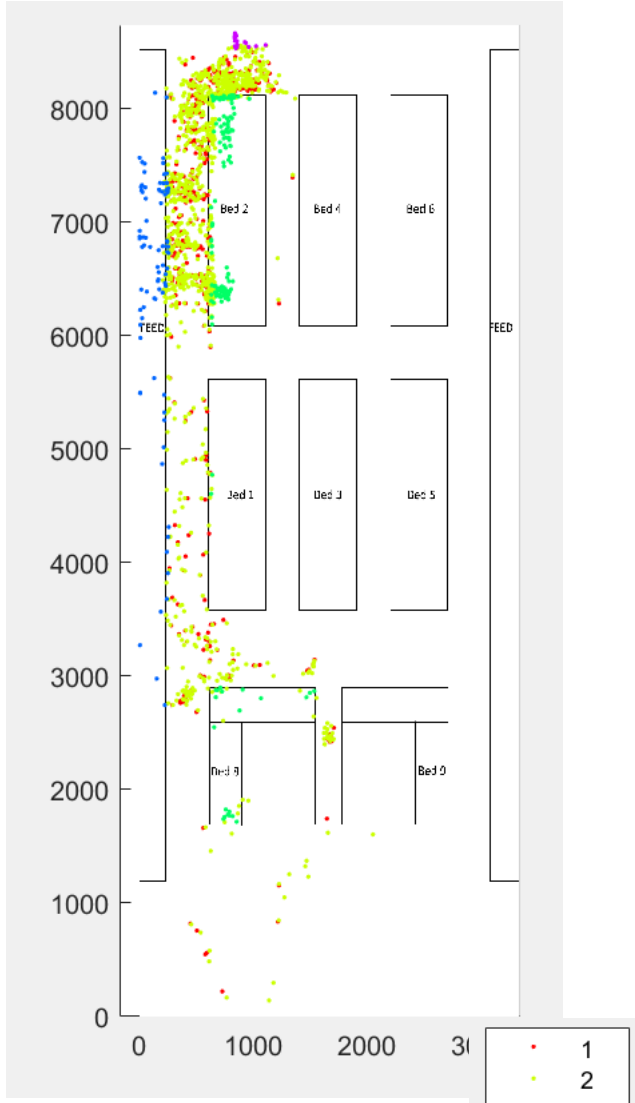
PA,2196869,00218585,1427847437212,1427847442656,3244,5717,137,2,8

PA,2196869,00218585,1427847442656,1427847486835,3216,5682,137,1,0

PA,2196869,00218585,1427847486835,1427847489255,3156,5713,137,2,25

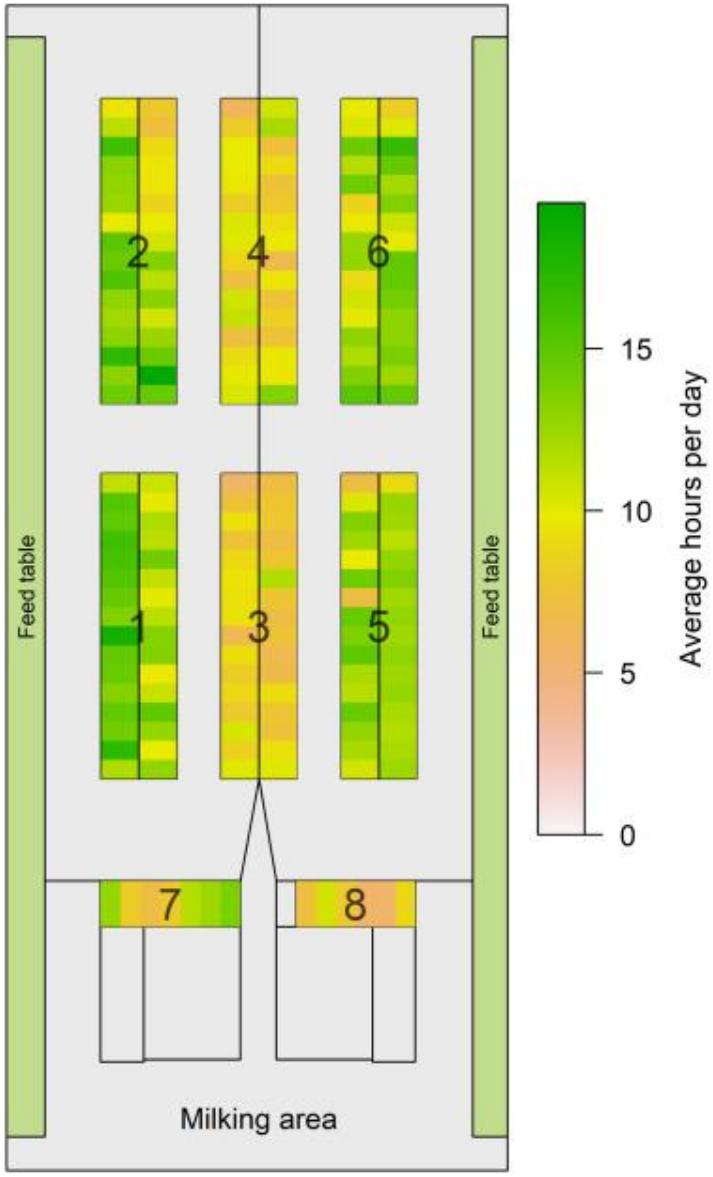
PA,2196869,00218585,1427847489255,1427847796661,3134,5735,137,1,0





1 Standing
2 Walking
3 Bed
4 Feeding
999 Out range

How we use PA data?
An example: Cubic Choice



Churakov, M., Silvera, A. M., Gussmann, M., & Nielsen, P. P. (2021). Parity and days in milk affect cubicle occupancy in dairy cows. *Applied Animal Behaviour Science*, 105494.

Gussmann, M., Marina, H., Ren, K., Rönnegård, L., & Nielsen, P. P. (2025). Variations in cow behaviour after regrouping in a conventional Swedish dairy herd. *Applied Animal Behaviour Science*, 106790.

PAA
(id, span, interval, activity type, duration, periods, distance)

data_entity,tag_id,tag_string,span,interval,activity_type,distance,periods,duration

PAA,2197312, 00218740,1381755600000,3600000,1,0,88,1204757

PAA,2197312, 00218740,1381755600000,3600000,2,2710,92,1352724

PAA,2197312, 00218740,1381755600000,3600000,3,0,3,555727

PAA,2197312, 00218740,1381755600000,3600000,4,0,2,486791

PAA,2197139, 00218693,1381755600000,3600000,1,0,29,209335

PAA,2197139, 00218693,1381755600000,3600000,2,3502,58,186696

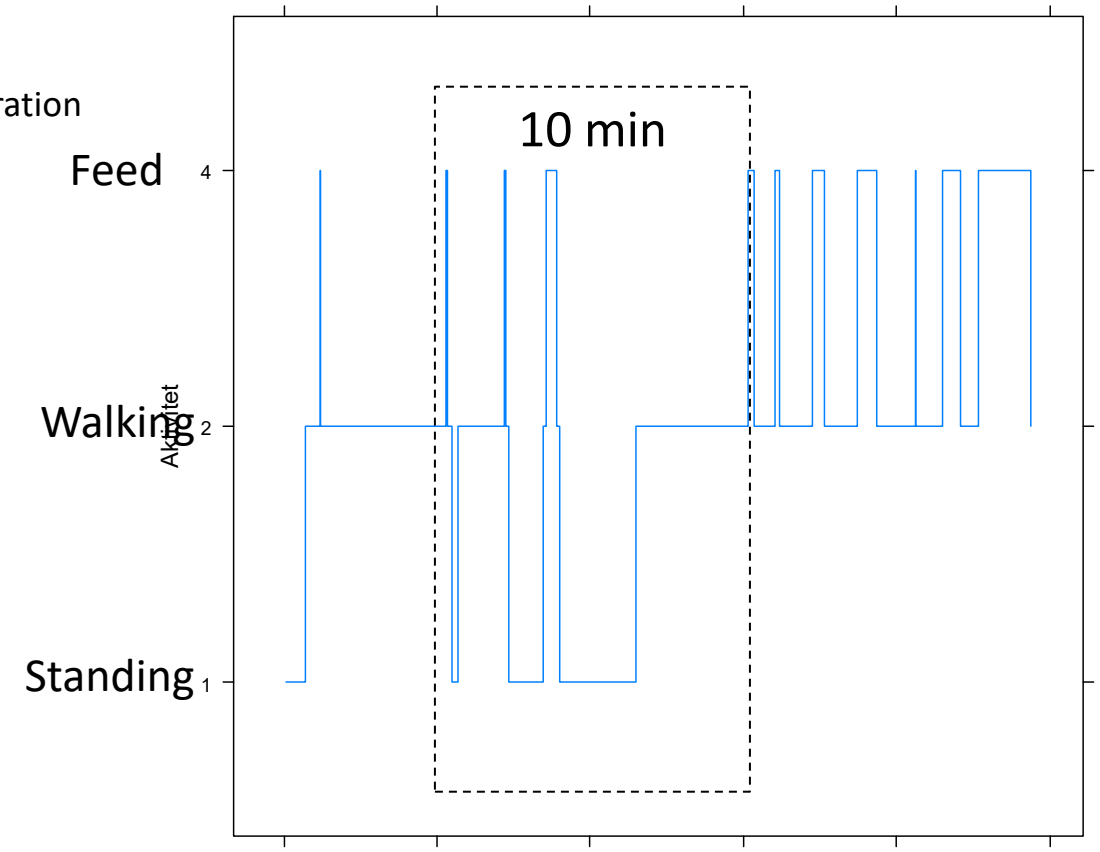
PAA,2197139, 00218693,1381755600000,3600000,3,0,29,3203968

PAA,2196892, 0021859C,1381755600000,3600000,1,0,19,194937

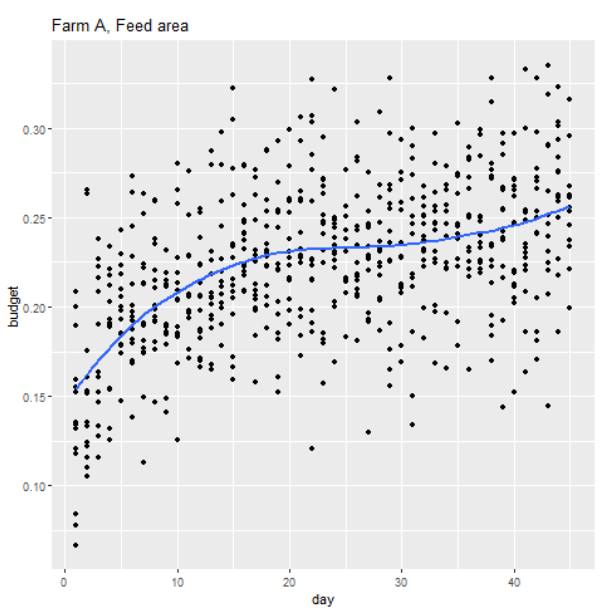
PAA,2196892, 0021859C,1381755600000,3600000,2,2055,30,388698

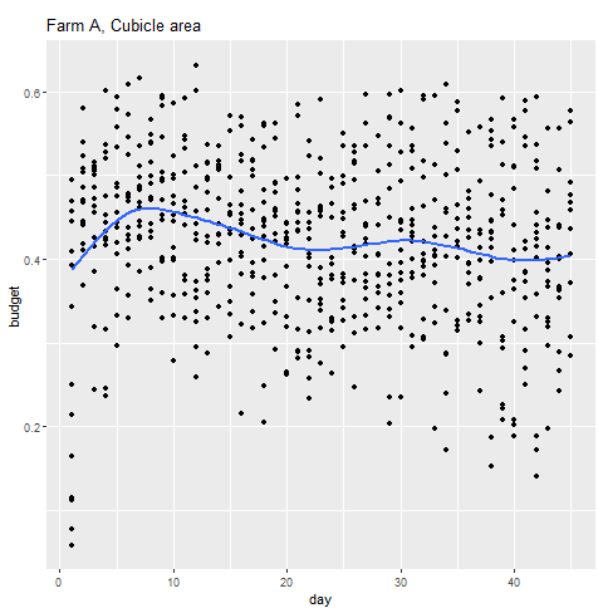
PAA,2196892, 0021859C,1381755600000,3600000,3,0,4,58709

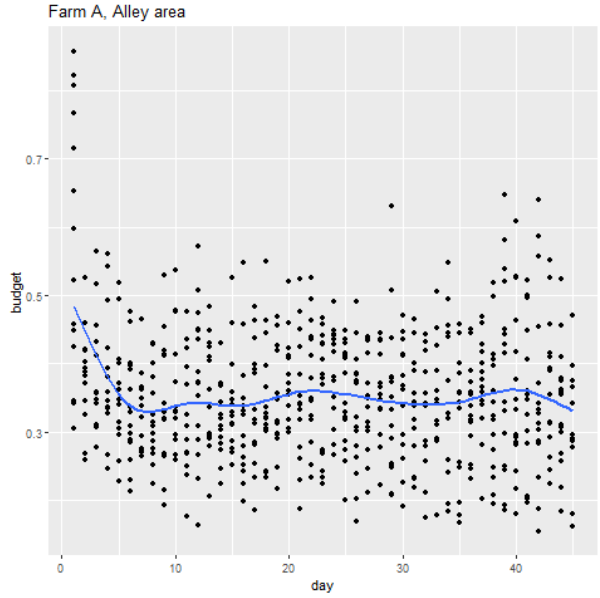
PAA,2196892, 0021859C,1381755600000,3600000,4,0,8,2957655



How we use PAA data?
An example: Activity budget







Gussmann, M., Marina, H., Ren, K., Rönnegård, L., & Nielsen, P. P. (2025). Variations in cow behaviour after regrouping in a conventional Swedish dairy herd. *Applied Animal Behaviour Science*, 106790.

Combine time budget to activity

Fourier-Based Approximation with Thresholding (FBAT)

Git: https://github.com/nicolas-wagner/FBAT

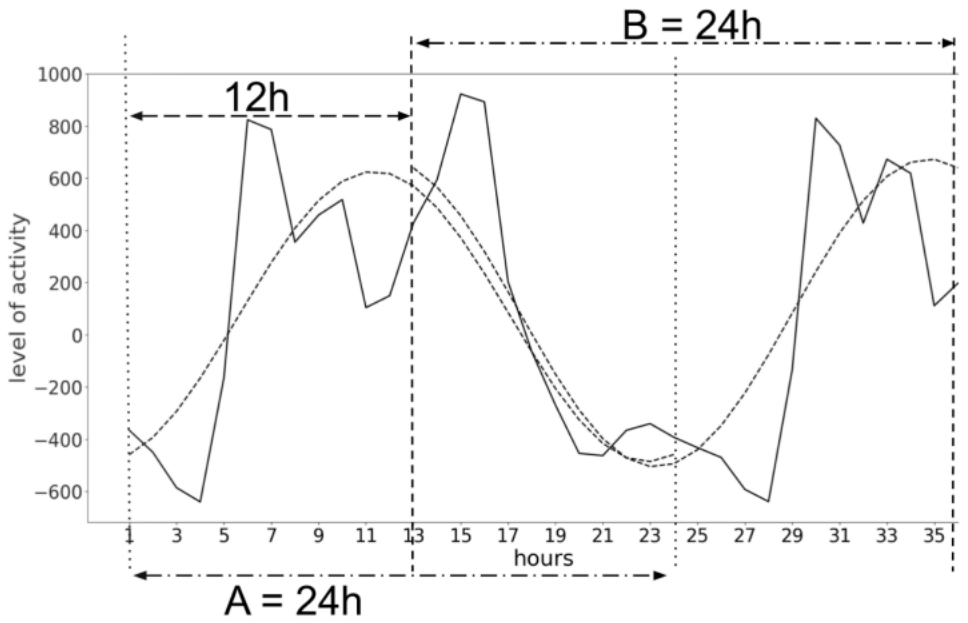


Fig. 1. Example of a 36-h time series of cow activity modelled with a Fourier transform. Solid line: activity level calculated from basic activities (weighted sum of the time spent 'resting', 'in alleys' or 'eating', unitless). Dotted lines: Fourier transform of the first and last 24-h segments of this 36-h time series.

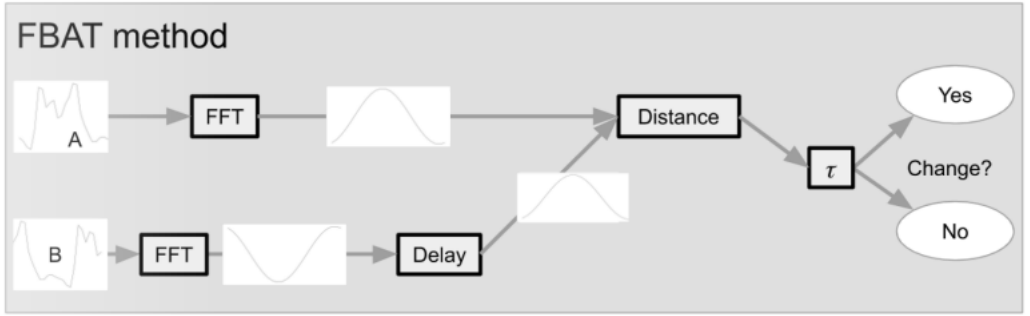


Fig. 2. Framework of the FBAT method to detect changes in circadian activity rhythm. Within a 36-h time series, we used Fast Fourier Transform (**FFT**) to model the variations in activity during the first and last 24-h segments of this 36-h series. After aligning the two models in time, we calculate the Euclidian distance between them and then compare that distance to a given threshold, above which we consider that the rhythm has changed.

Wagner, Nicolas, et al. "Detection of changes in the circadian rhythm of cattle in relation to disease, stress, and reproductive events." *Methods* 186 (2021): 14-21.

Exercise time

A few questions after you finished go through the script:

- From the *analysis. R* you have the *getMeanPos* function to get the mean position of an individual (or a group). You can compare one individual's mean position with the cubic preference.
- Compare the cubic preference differences between Mid-lactation &&Parity 1 and Mid-lactation Parity 3+.
- Get the average time each cubicle was occupied of the day.