A Hybrid CUSUM approach to identify residency and transition periods for animal movement with an application to housed dairy cows

Alexander Kieran Partner

1606829

Supervisor: Professor E A Codling

A thesis presented for the degree of

Master of Science (by Dissertation)

Department of Mathematics

University of Essex

March 2018

Contents

Abstract

Over recent years there has been a significant advance in tracking technology which has allowed animal movement data to be collected with greater accuracy and in larger quantities. As a result, there has been a need for developing techniques to analyse this data and gather information such as bouts of foraging behaviour in wild animals, understanding how an animal's movements relate to the resources within an environment or to provide potential indicators of animal welfare. The aim of this report is to develop an automated behavioural classification for residency and transition periods based on two-dimensional position data via a Hybrid CUSUM method. It is hoped that by detecting the amount of movement of an individual over a selected period of time and comparing it with expected results of healthy individuals will lead to an indication of welfare although this CUSUM method can also be used for other purposes such as gathering information about foraging patterns. The versatility of this method is that it can be applied to any animal that exhibits mostly residency and transitory behaviour. In order to automatically identify the residency and transition periods, a Hybrid CUSUM method has been developed and has been applied to a data set involving housed dairy cows in the hope to identify differences in the typical movements of lame cows against the typical movements of non-lame cows. The Hybrid CUSUM has the novelty that the standard deviation is predefined to correspond to the expected deviation of an animal whilst in a state of residency, whilst the mean is calculated directly from the data. It also has the novelty that a second algorithm is initiated once the system is in a state of out-of-control to detect when the system is back in-control and then the CUSUM will restart around a new

mean value. The outcome of this report is the ability to automatically identify when cows are resident within three distinct areas of the barn (resting area, feeding area and milking area) and to imply the types of transitions between areas of the barn. In earlier work, attempts to classify the behavioural state of cows using the same data set have been achieved using accelerometer data, whereas this report uses position data to identify residency periods within a particular zone of the barn and infer types of transitions between zones such as resting to feeding or transitions within the same zone such as feeding to feeding. It is found that non-lame cows tend to move around more and spend longer total time within the feeding area than lame cows.

1 Introduction

1.1 Motivation

The aim of this report is to develop a Hybrid CUSUM approach that can automatically identify periods when animals are stationary and periods when they are moving - considered as residencies and transitions respectively - in order to give welfare indicators of animals based solely on their movement behaviour. In other applications, using a Hybrid CUSUM to detect residency and transition behaviour may give indicators of the search strategies of an animal or offer insights about the resources within an environment which an animal lives in. This CUSUM approach has been applied to a data set involving housed dairy cows since dairy cows exhibit residency and transition behaviour, partly to verify that the CUSUM method works as expected but also to analyse potential differences in the movement behaviour of lame cows compared to that of non-lame cows. By comparing the differences in movement behaviour between these two groups of cows, it is hoped that these insights can be developed within the dairy industry to give indicators of possible lameness within a cow and enable faster treatment.

1.2 Previous Work

A similar form of CUSUM applied to the same cow data set has been developed simultaneously by another student Woor (2017) and, as such, the work outlined in this report provides a form of validation whilst also being an extension to this method. All work has been carried out totally independently so as to verify the validity of both the methodology and the results. The Hybrid CUSUM described in Woor (2017) uses a *fixed* origin to measure distance, whereas the CUSUM described in this report differs so that the origin *moves* according to the location of the most recent residency (see Section 3.4). The novelty of the shifting origin provides enough of an extension to differ from previous work and makes the method general enough to be applicable to a multitude of other animal movement data sets, whereas the Hybrid CUSUM method with the fixed origin in Woor (2017) is limited to this particular cow data set or data sets where the animal movements are suitably restricted by geographical layout.

2 Background

2.1 Animal Tracking

Over recent years there has been a significant increase in the amount of tracking projects of animal species in an attempt to provide an indicator of behavioural patterns and welfare indicators (Nathan et al., 2008). Typically animal movement data is collected as either accelerometer signal (Moreau et al., 2009; Robert et al., 2009) which records the activity of the animal through time or as position data (either two-dimensional or three-dimensional) (Tomkiewicz et al., 2010; Kays et al., 2015) which tracks the location of the animal through time. As a consequence of these advances in technology, the demand for accurate and efficient methods to process and interpret the vast amount of data collected from these projects has become more necessary and will continue to do so as the tracking technology continues to improve (Voegeli et al., 2001; López-López, 2016; Jukan et al., 2017).

2.2 Existing Methods for Animal Movement

2.2.1 Hidden Markov Models

Recent studies have used established techniques such as Hidden Markov Models (HMM) (Franke et al., 2006; Jonsen et al., 2007; Patterson et al., 2008; Langrock et al., 2012; Woillez et al., 2016) which attempts to derive a walker's hidden state sequence from a discrete sequence of observed variables based on movement quantities such as step lengths and turning angles. The hidden state sequence is typically behavioural states such as “encamped” or “exploratory” (Morales et al., 2004) which roughly corresponds to an animal either staying within a particular location seen to be its regular point of residence or moving in a state of exploration but other biological states may be directed towards survival states and extensions can be made for environmental factors (Morales et al., 2004). HMM may also be used to extract location of animals from noisy GPS data where the discrete nature of the data collection provides an ideal candidate for a HMM to be applied. HMM are composed of two probabilistic models which change over time: the transition model which describes the sequence of states over time; and the observation model which describes the probability of seeing another state given any currently observed state. It is assumed that the Markov property holds: that is, the probability of seeing a state at time t_n depends only on the state at t_{n-1}. In the context of animal movement data, the transition state model may correspond to the position, speed or direction variables of a walker and is usually the hidden state sequence whilst the observed sequence may be environmental factors such as temperature. For example, in Woillez et al. (2016) the intention was to analyse the location of pelagic fish. Here the hidden state sequence was two-dimensional graphical space whilst the observed sequence was data collected from sensors involving temperature and depth.

2.2.2 Area Restricted Search

Area Restricted Search (Fauchald & Tveraa, 2003; Weimerskirch et al., 2007; Knell & Codling, 2012) refers to the foraging pattern of species when attempting to find resources such as food and as a result it is a highly useful model for establishing animal search strategies. It is typically the case that walkers slow down once they are in a resource-rich environment and will stay there for longer. A forager typically exhibits two types of strategies: the first is an **intensive** search which can be seen through small step lengths and high turning angles which means that an area is searched in high detail but at a slow rate; the second is an **extensive** search which can be seen through large step lengths and low turning angles which means that an area is not searched in high detail but is covered at a fast rate. As a result, the transitions between intensive and extensive search strategies is highly relevant not only for identifying search strategies of foragers but also for gathering information about the quantity of resources within an environment. A statistical approach that can be used is a Moving Average (Robinson et al., 2007) which states that the localised net average velocity or turning angles between an animal's movement will be different between intensive and extensive search strategies. This can then be used to detect the type of search strategy that a walker is performing. Another analysis approach that can be used is First Passage Time (Tuckwell & Wan, 1984) which refers to the amount of time that an animal spends within a particular location, typically seen by overlaying a circle over a plot of the animal's position. An intensive search strategy should mean that an animal remains in the circle for a longer period of time, whereas an extensive search strategy should mean that the walker clears the circle relatively quickly. It was proposed in Fauchald & Tveraa (2003) that the optimal circle size should have a radius roughly equal to the diameter of the size of a typical food patch for that animal, however Barraquand & Benhamou (2008) argued that there was no evidence to support this hypothesis. Fractal Dimension (Dicke & Burrough, 1988) in movement analysis refers to the fractal nature of an animal's path. It is assumed that an intensive search strategy should display a larger fractal dimension since an animal remains in a location for a longer period of time, whereas an extensive search strategy should display a smaller fractal dimension. Change Point Detection refers to techniques that attempt to detect significant changes in a walker's path by plotting the path signal over time. Such methods used to detect changes include Line Simplification (Douglas & Peucker, 1973; Visvalingam & Whyatt, 1992; Thiebault & Tremblay, 2013) which looks to reduce the number of vertices in a trajectory to monitor the impact of path topology and hence determine change points; Change Point Test (Byrne et al., 2009) which detects significant changes in the orientation of a walker between two points of a trajectory; Piecewise Regression (Buchin et al., 2013) which partitions a path into segments and looks to fit a regression line in each segment to determine significant gradient changes between segments; and Pruned Exact Time (PELT) Algorithm (Killick et al., 2012).

2.2.3 Levy Flights

Another established theoretical model that is commonly used for animal movement is Levy Flights (Viswanathan et al., 2000; Codling & Plank, 2011; Zaburdaev et al., 2015) which hypothesises that animals follow a random walk with step lengths from a probability distribution with heavy tails. Levy Flights are related to the context of this report since the movement patterns produced are similar to that of alternating extensive and intensive search strategies seen in Area Restricted Search behaviour

(Section 2.2.2). However, there is still debate as to whether Levy Flight models are truly optimal search strategies in patchy environments since classical random walks such as composite correlated random walks can be more efficient as a search strategy that an animal might use (Benhamou, 2007). Following on from Levy Flights is the Levy Flight Foraging Hypothesis (Sims et al., 2008; Humphries et al., 2010; James et al., 2011; Sims et al., 2012) which proposes that when predators are struggling to find food, they abandon the type of random walk typically defined as Brownian Motion in favour of the rapid, sharp movements and long trajectories of a Levy Flight which enables very fast and efficient diffusion. This type of behaviour is often described as fractal movement. An example of where this movement has been explicitly observed is in marine predators (Travis, 2008; Reynolds, 2014). It has been shown that animals that display this type of foraging behaviour have a higher number of encounters with prey (Sims et al., 2008) and may suggest a predator strategy close to theoretical optimum. Despite this, research in recent years has suggested caution to Levy Flights and Walks due to their omission of biological realism (Benhamou, 2007; Pyke, 2015). In Pyke (2015) assumptions required for Levy Flights such as a featureless environment and lack of memory was discussed since this implied that animals cannot respond to past and present information which has questionable biological adequacy. Also questioned was the assumption that animals are searching for food or some other resource whereas this may not always be the case; Sueur (2011) suggested that animals were just moving between known locations, thus having knowledge of the surrounding environment violating two assumptions of the Levy Flight. There has also been concern raised with the assumptions of the abundance and density of food resources within an environment and whether the Levy Walk model is truly optimal for foraging. Initially it was suggested that when food is “non-depleting” and occurs at low density (Viswanathan et al., 1999) a Levy Walk can be used but then others suggested it was suitable for when resources were sparse (Humphries et al., 2012). As such other possible models may offer superiority to the Levy Walk (Hills et al., 2013).

2.2.4 Home Range Analysis

Home Range Analysis (Burt, 1943; Jennrich & Turner, 1969; Ford & Krumme, 1979; Aebischer et al., 1993) refers to the area where an animal lives and moves on a periodic basis. Home ranges may then be able to infer information about resources within an environment or general animal behaviour based on the size and shape of the home range (Van Winkle, 1975). It is typically the case that home ranges display Area Restricted Search strategies associated with intensive searches (Section 2.2.2). One such method to determine the home-range is to use a kernel-density method (Worton, 1989; Seaman & Powell, 1996) which aims to estimate the distribution of an animal's position (utilisation distribution) by means of a probability density function placed over each point in the location data set corresponding to the relative amount of time that an animal spends in any one place (Van Winkle, 1975). Another method used to estimate home range is to construct the minimum convex polygon (MCP) which completely contains the home range and is still widely in use (Baker, 2001). Whilst the kernel-density method and the MCP are amongst the more established methods, more recent developments have shown that a Brownian Bridge Movement

Model (BBMM) (Horne et al., 2007) and a Dynamic BBMM (Bullard, 1999; Calenge, 2006) may offer a suitable model for analysing Home Ranges by considering that the probability of an animal being in an area is conditioned on the start and end locations, the elapsed time between points and the speed of movement. Another recent method is the Time Geographic Density Estimation (TGDE) (J. A. Downs et al., 2011; J. Downs et al., 2018) which generates a continuous probability distribution of an animal's location over time from tracking data, thus producing contours of relative intensity. TGDE is a favoured method since it is based on the animal's speed of movement and so does not create "false intensities" where an animal could not have been due to the location and time constraints from the data. There has also been scope to explore the physical overlap of home ranges of multiple animals and explore how the dynamics and behaviour of these animals change based on interactions (Macdonald et al., 1980).

3 Hybrid CUSUM Methodology

3.1 Outline of a CUSUM

The aim of this report is to develop a method that uses two-dimensional position data to detect residency and transition periods within animal movement. This method can then be applied generally to animal movement data in an attempt to identify welfare from movement behaviour.

The concept of a cumulative sum (CUSUM) was introduced by E.S. Page (Page, 1954) as a statistical analysis tool to determine change detection. It has since been used in a wide variety of situations from the quality-control check of a machine manufacturing goods (Black & Mejabi, 1995) to ensuring consistent performance of anaesthetists (Dexter et al., 2014). It is designed to account for natural variability and only flag up when a situation falls consecutively below a standard specified by a tolerance level. The purpose of the CUSUM in this situation is to apply it to an animal data set in order to identify periods when an individual is staying relatively stationary (defined as a residency) and the periods when an individual is moving from one area to another (defined as a transition).

3.2 Decision-Interval CUSUM

The Decision-Interval CUSUM (DI-CUSUM) method (Hawkins, 1987; Montgomery & Gerth, 1998) is a signal analysis tool used to detect change in a system whilst having the ability to overcome noise. It can be used to determine situations such as the state of a manufacturing process (or the movement state of individuals as outlined here) as either “in-control” or “out-of-control” and determines whether relevant corrective action needs to take place. The method works by forming a cumulative sum of standardised observation values to determine whether they have deviated significantly from the reference mean. In order to compute the i^{th} indicator value Z_i the observation values X_i are standardised according to the reference mean $\overline{X}$ and the control standard deviation σ:

$$Z_i = \frac{X_i - \overline{X}}{\sigma}$$

In a DI-CUSUM, *the reference mean and standard deviation are predefined before the CUSUM process takes place* and remain fixed at these values throughout the process.

Since there will be indicator values above the control mean ($Z_i > 0$) and also indicator values below the control mean ($Z_i < 0$), two separate CUSUMs are computed for each case (Hawkins & Olwell, 1998), termed "Upper CUSUM" (θ^+) and "Lower CUSUM" (θ^-) respectively. These are both initially set to 0:

$$\theta_0^+ = 0 \quad \text{and} \quad \theta_0^- = 0$$

From this, the CUSUM values θ_i compare whether the observation values are showing a tendency to deviate by summing the previous CUSUM values θ_{i-1} to the current indicator value Z_i by testing whether the result of the sum is within a tolerance level h. In order to ensure that the upper CUSUM remains non-negative and the lower CUSUM remains non-positive, a *maximum* and *minimum* check with 0 is used respectively:

$$\theta_i^+ = \max\{0, \theta_{i-1}^+ + Z_i - k\} \quad \text{and} \quad \theta_i^- = \min\{0, \theta_{i-1}^- + Z_i + k\}$$

To determine whether the result of the CUSUM is significantly large or relatively small, a further tolerance level is set by using the control parameter h. The system is in-control when the CUSUM values are within the tolerance level:

$$\theta_i^+ < h \quad \text{or} \quad |\theta_i^-| < h$$

The system is out-of-control when the CUSUM values are outside the tolerance level:

$$\theta_i^+ > h \quad \text{or} \quad |\theta_i^-| > h$$

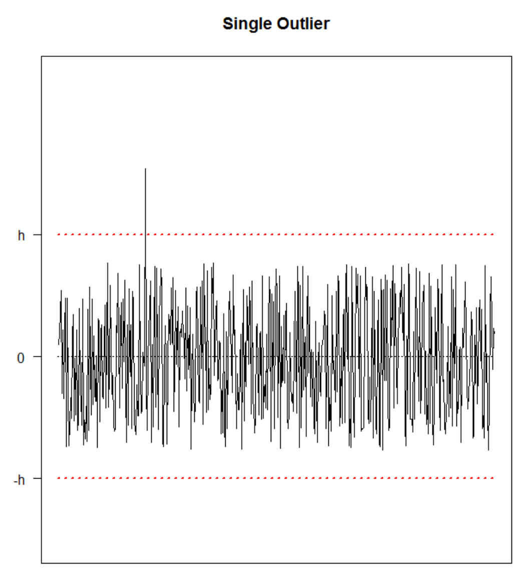
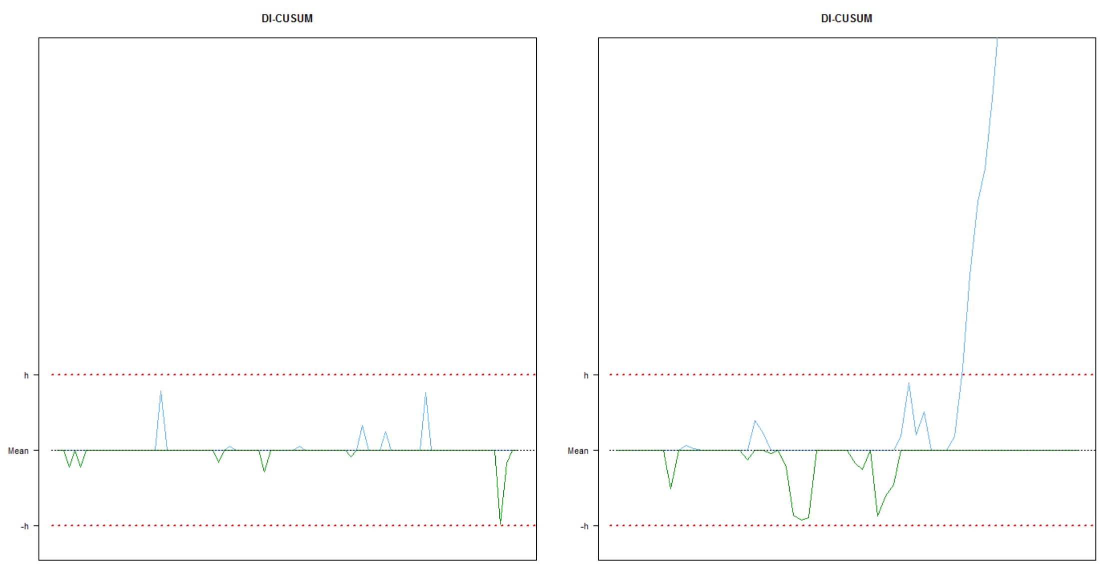


Figure 1: An illustration of a set of observation values with a single value being outside the tolerance level set by h. Using a system which merely considers whether observation values are above or below the tolerance level, this outlier would flag the system as out-of-control.

The difference between a CUSUM method compared to purely testing whether each individual indicator value is within a particular tolerance level (Figure 1) is that it is the *trend* of whether values are starting to deviate away from a reference mean that is being analysed. This means that individual outliers may not put the system into a state of out-of-control (depending on the leniency of the allowance parameter k) unless consecutive indicator values are consistently starting to deviate from the reference mean (Figure 2).



(a) In-control CUSUM

(b) Out-of-control CUSUM

Figure 2: An illustration of an in-control (Figure 2a) vs out-of-control (Figure 2b) DI-CUSUM signal. CUSUM values above the mean line are termed "Upper CUSUM" and are displayed in blue and below the mean line are termed "Lower CUSUM" and are displayed in green. When the CUSUM values are inside the tolerance level set by the value h the system is said to be in-control, whereas as soon as the CUSUM values go outside the tolerance level the system is said to be out-of-control and the user stops the process immediately after the DI-CUSUM is seen to go out-of-control.

3.3 Self-Starting CUSUM

The DI-CUSUM is applicable to the quality control check of processes where the expected mean and standard deviation are known beforehand such as a machine manufacturing goods required to be of a certain size or weight. However, certain situations show less concern for the system to be constrained between strict values and direct more attention towards the broader question of whether the system is exhibiting the behaviour of being in-control or out-of-control. As such, these situations require *the mean and standard deviation to be calculated directly from the observation values* whilst the CUSUM is running. This refinement of the DI-CUSUM is called a Self-Starting CUSUM (SS-CUSUM) method (Hawkins, 1987).

A type of SS-CUSUM would be better-suited to the application outlined in this report than a DI-CUSUM since the aim is to detect animal deviation from particular locations in order to identify residencies and transitions. Clearly, the residency positions will occur at different locations and as such require a mean value to be calculated directly from the data to give the reference mean rather than being predefined seen with the DI-CUSUM.

The mean $\overline{X}$ and standard deviation σ are constructed by using consecutive values up to the current ith position:

$$\overline{X_i} = \frac{\sum_{k=1}^{i} X_k}{i} \quad \text{and} \quad \sigma_i = \sqrt{\frac{\sum_{k=1}^{i}(X_k - \overline{X_i})^2}{i}}$$

meaning that these values will become increasingly accurate as more values are considered. As a logical result of this, the SS-CUSUM is initially sensitive to noise and early outliers which may put the system in an out-of-control state too early. An approach which can be useful for dealing with early outliers is metric winsorization (Hawkins & Olwell, 2012; Pazhayamadom et al., 2015) which essentially edits the larger deviation due to the outlying values to more central values by applying a cut-off threshold known as the winsorizing constant w. The upper and lower CUSUMs are also updated using this winsorizing constant. Although useful in certain contexts, metric winsorization is not overly useful for the application of a CUSUM considered in this report since the the aim is to detect residency and transition states from in-control and out-of-control states respectively (Section 3.4). It is not useful here since the small level of error in the system is overcome by using an adequate value for the standard deviation (Section 3.5.1) and as such, any large outliers are probably due to the animal moving (transitory) which means that the system should be detected as out-of-control.

Using these values of the mean and standard deviation, the CUSUM is constructed in the same way as the DI-CUSUM, with standardised observation values X_i forming the indicator values Z_i:

$$Z_i = \frac{X_i - \overline{X_i}}{\sigma_i}$$

The Upper and Lower CUSUM are both initially set to zero

$$\theta_0^+ = 0 \quad \text{and} \quad \theta_0^- = 0$$

and then defined as

$$\theta_i^+ = \max\{0, \theta_{i-1}^+ + Z_i - k\} \quad \text{and} \quad \theta_i^- = \min\{0, \theta_{i-1}^- + Z_i + k\}$$

where k is the allowance parameter. The tolerance level for the CUSUM is set by h, with the system being in-control when the CUSUM values are within the tolerance level:

$$\theta_i^+ < h \quad \text{or} \quad |\theta_i^-| < h$$

and out-of-control when the CUSUM values are outside the tolerance level:

$$\theta_i^+ > h \quad \text{or} \quad |\theta_i^-| > h$$

3.4 Hybrid CUSUM

The CUSUM approach outlined in this report is a hybrid in two respects: firstly, whilst the mean is calculated from observation data as seen in a SS-CUSUM (Section 3.3), the standard deviation is predefined (Section 3.5.1) as seen in a DI-CUSUM (Section 3.2); secondly, once the CUSUM detects the system to be in a state of out-of-control, a second algorithm is implemented to detect when the system is considered to be back in-control (Section 3.5.2) and the CUSUM restarts from when the system is thought to be back in-control (Figure 3).

For a Hybrid CUSUM, the system is initially defined as being in-control. The standard deviation σ is predefined and the mean $\overline{X}$ is calculated by using consecutive values up to the ith position:

$$\overline{X_i} = \frac{\sum_{k=1}^{i} X_k}{i}$$

For as long as the CUSUM is defined as being in-control, the mean is calculated in this way, meaning that the longer the CUSUM is in-control for, the more-accurate the mean value becomes since more values are being considered. Each iteration of the in-control CUSUM is standardised to produce an indicator variable Z_i:

$$Z_i = \frac{X_i - \overline{X}}{\sigma}$$

where X_i is the i$^{\text{th}}$ value being considered.

The Hybrid CUSUM is otherwise constructed in a similar way to that of the DI-CUSUM or SS-CUSUM with Upper and Lower CUSUM values initially set to zero:

$$\theta_0^+ = 0 \quad \text{and} \quad \theta_0^- = 0$$

then defined as:

$$\theta_i^+ = \max\{0, \theta_{i-1}^+ + Z_i - k\} \quad \text{and} \quad \theta_i^- = \min\{0, \theta_{i-1}^- + Z_i + k\}$$

where k is an allowance parameter. The Hybrid CUSUM is defined as out-of-control when either the θ^+ or θ^- exceeds the tolerance value h:

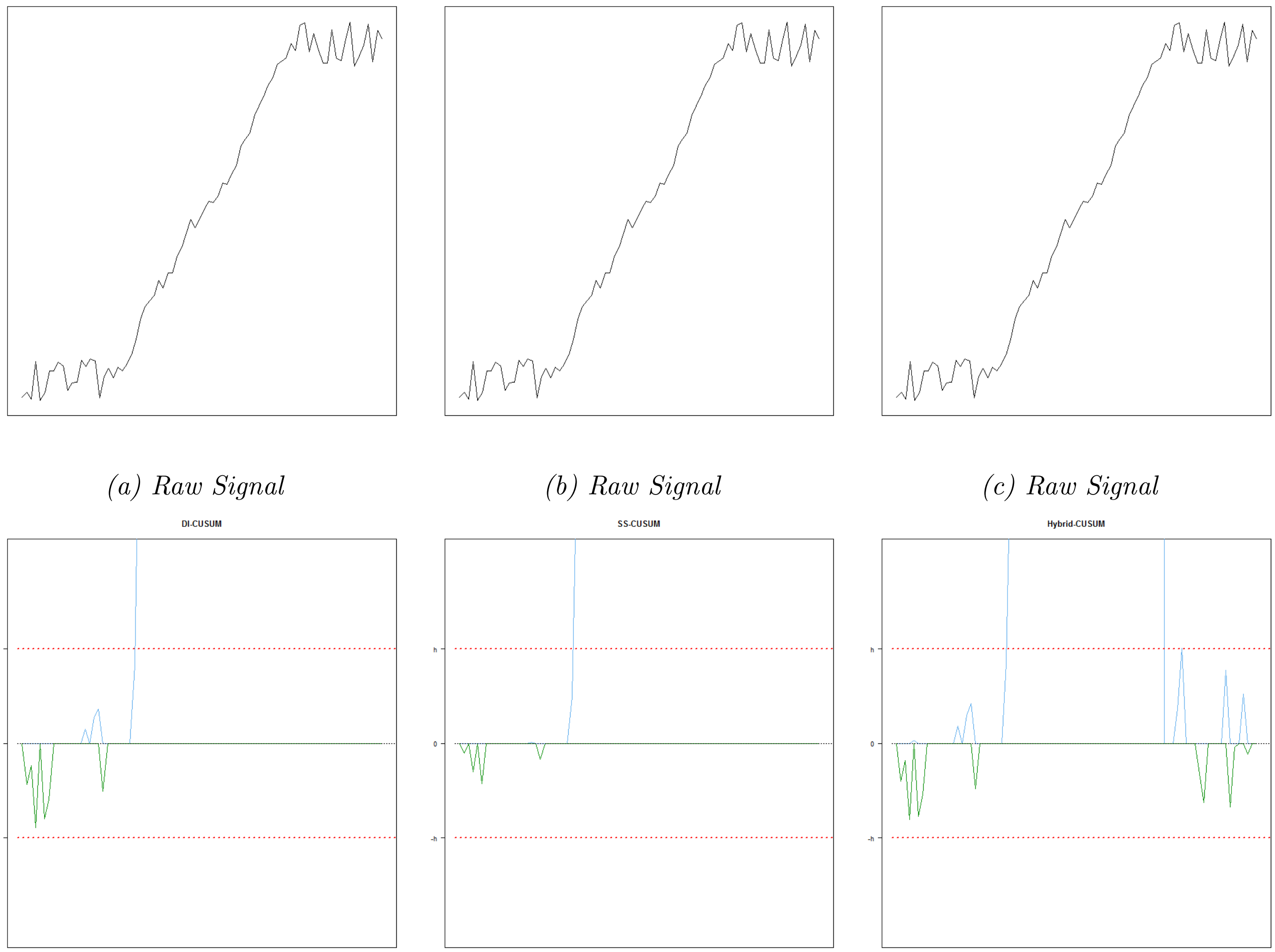
$$\theta_i^+ > h \quad \text{or} \quad |\theta_i^-| > h$$

Whereas a DI-CUSUM or SS-CUSUM is set to stop once the system is detected as being in a state of out-of-control, one of the novelties of the Hybrid CUSUM is to initiate a second algorithm to detect when the system comes back in-control and then automatically restart the Hybrid CUSUM process from this new in-control point (Section 3.5.2). Once the system is defined as in-control at position j, the Hybrid CUSUM runs using a similar method to that outlined above, with the same predefined standard deviation and the mean calculated from the jth position up to the ith position for which the CUSUM is still defined to be in-control:

$$\overline{X_i} = \frac{\sum_{k=j}^{i} X_k}{i-j}$$

and Upper and Lower CUSUM values at position j set to zero:

$$\theta_j^+ = 0 \quad \text{and} \quad \theta_j^- = 0$$



(a) Raw Signal *(b) Raw Signal* *(c) Raw Signal*

(d) DI-CUSUM *(e) SS-CUSUM* *(f) Hybrid CUSUM*

Figure 3: A comparison of how a DI-CUSUM (Figure 3d), SS-CUSUM (Figure 3e) and Hybrid CUSUM (Figure 3f) may deal with the same type of raw signal (Figures 3a, 3b and 3c). The Upper CUSUM θ^+ is drawn in blue and the Lower CUSUM θ^- is drawn in green. The mean is marked on the CUSUM plots and the red dotted line corresponds to the tolerance level h. The system is defined as being out-of-control when either θ^+ or $|\theta^-|$ exceeds h.

3.5 Applying the Hybrid CUSUM to animal position data

Since this application requires the CUSUM to detect an animal changing locations, two-dimensional location data is converted into one-dimensional distance data by measuring the distance of the individual from a given origin position:

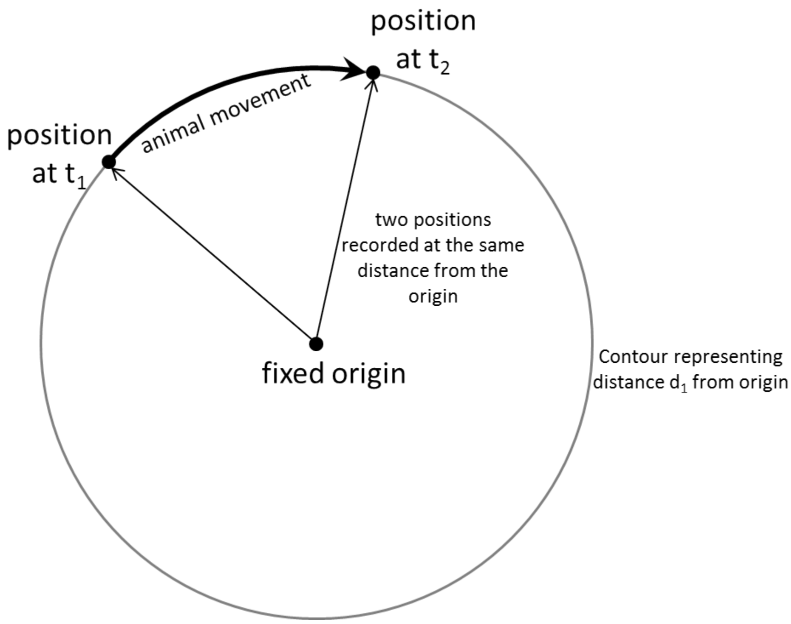
$$\text{Distance From Origin} = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}$$

where (x_0, y_0) is the location of the origin and (x_i, y_i) is the location of the animal at a time step i.

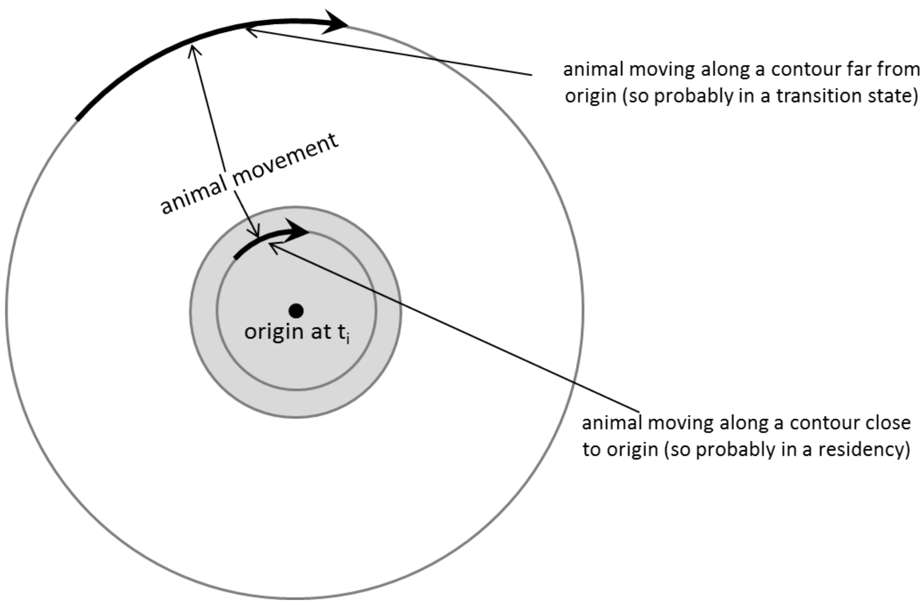
Aside from fixing the standard deviation whilst calculating the mean from the data and allowing the CUSUM to restart once the system is detected as back in-control (Section 3.4), another of the novelties of this particular application of a CUSUM is to implement a shifting origin so that the distance from the origin (DFO) is always calculated from the last-known residency position rather than from an arbitrary fixed location. If distance is measured from an arbitrary fixed origin position then this could theoretically result in a transition movement along a contour not being detected because the animal would be regarded as a consistent distance from the origin whilst on that contour (Figure 4a). This approach of using a fixed origin was outlined in Woor (2017) where a similar Hybrid CUSUM was applied to the same cow data set described in this report. Whilst this application could not easily be applied to other animal movement data sets due to the disadvantages already described, it worked for the particular application in Woor (2017) since the data set in question involved dairy cows whose movements were restricted by the layout of the barn which generally prevented cows from moving along contours for considerable periods of time. The CUSUM involving a shifting origin described in this report is a refinement of this fixed-origin CUSUM and has the advantage that it can be applied more generally to other animal movement data sets by continuously moving the origin position to the most-recent residency position. As such, even if an animal is moving along a contour, the CUSUM will only detect a transition if the individual has deviated far enough from the last-known residency position (Figure 4b).

When the Hybrid CUSUM is in-control the animal is defined to be in a state of *residency* since the location of the individual remains suitably close together. By contrast, when the Hybrid CUSUM is out-of-control the animal is defined to be in a state of *transition* since the location of the individual is shown to be moving away from previous positions. For convenience, the animal is initially defined to be in-control and the initial origin position (x_0, y_0) is calculated by using the mean of the first ten position coordinates of the animal. The Hybrid CUSUM then runs in an identical way to that described in Section 3.4 with the mean of DFO values being used to form the running mean $\overline{X}$. Considering the data has been standardised, a k and h value of 1 is deemed to be appropriate since this corresponds to one standard deviation or an allowable deviation within a circle of radius three metres (see Section 3.5.1).

Once an animal has entered a transition state, the system should enter a state of out-of-control and the mean of all the x and y ordinates whilst the system has most-recently been in-control is recorded as the best approximation of the most-recent residency position and a second algorithm (see Section 3.5.2) is initiated which detects when the system comes back in-control. When the system re-enters a state of



(a) Fixed origin



(b) Shifting origin

Figure 4: An illustration of the limitations of a fixed origin position and the advantages of a shifting origin position.

in-control, this should correspond to an animal entering a residency state and the Hybrid CUSUM restarts from this point using the same method identified in Section 3.4.

3.5.1 Standard deviation

Fixing the value for standard deviation can be justified from a biological perspective since, in this context, it is the expected amount of deviation that an animal is expected to make whilst in a state of residency. For instance, it is highly unlikely that once an animal is classified as being in a residency it will remain completely static in that location without moving around slightly. Clearly, if the standard deviation is too small then even slight movements would most-likely get detected by the CUSUM as being out-of-control and hence flag the animal as being in a state of transition, whereas an excessively large value may result in key transitions being missed due to being included within the allowable deviation from the residency position. This idea was developed further for the application to dairy cows considered in this report and is discussed in section 4.1.1.

3.5.2 Detecting when the system is back in-control

Once the CUSUM moves to an out-of-control state, the Hybrid CUSUM is stopped and a second algorithm is initiated to determine when the system is back in-control. This is achieved by looking at a window size of n positions and analyses whether the position points within this window are clustered together or more spread-out (Figure

5). If a high number of the position points within this window are clustered together then this likely indicates that the animal has entered a state of residency and so the CUSUM algorithm can restart centered around this new residency position. If, however, the position points are more spread-out then this is likely because the animal is still moving and so the algorithm advances the window forward and continues to search for position points clustered together.

The working of this algorithm involves taking the mean of the next n position values from where the system is defined to be out-of-control at point i and this value becomes the origin (x_0, y_0):

$$(x_0, y_0) = \left(\frac{\sum_{k=i}^{i+n} x_k}{n}, \frac{\sum_{k=i}^{i+n} y_k}{n} \right)$$

where x_k and y_k are the kth position ordinates and i is the start value of the window.

The next n distances D are then measured from this origin position (x_0, y_0) using the respective position coordinates (x, y):

$$\begin{aligned}
D_i &= \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \\
D_{i+1} &= \sqrt{(x_{i+1} - x_0)^2 + (y_{i+1} - y_0)^2} \\
&\vdots \\
D_{i+n} &= \sqrt{(x_{i+n} - x_0)^2 + (y_{i+n} - y_0)^2}
\end{aligned}$$

Successive distances within this window of length n are standardised and compared to see whether they fall within a given tolerance level:

$$\frac{D_{i+1} - D_i}{\sigma} < t$$
$$\frac{D_{i+2} - D_{i+1}}{\sigma} < t$$
$$\vdots$$
$$\frac{D_n - D_{n-1}}{\sigma} < t$$

where t is a defined tolerance that the distances have to be within in order to suggest they are suitably close together.

If a defined number m of these successive distances are close together then the system can be considered as in-control and the CUSUM restarts around the central value of this window. If this condition is not met then the system is still defined to be out-of-control and the window of n distances to be tested is advanced to start at the $(i+1)$th position and the process is repeated. This algorithm continues until such a time when the successive distances are suitably close together to define the system as back in-control.

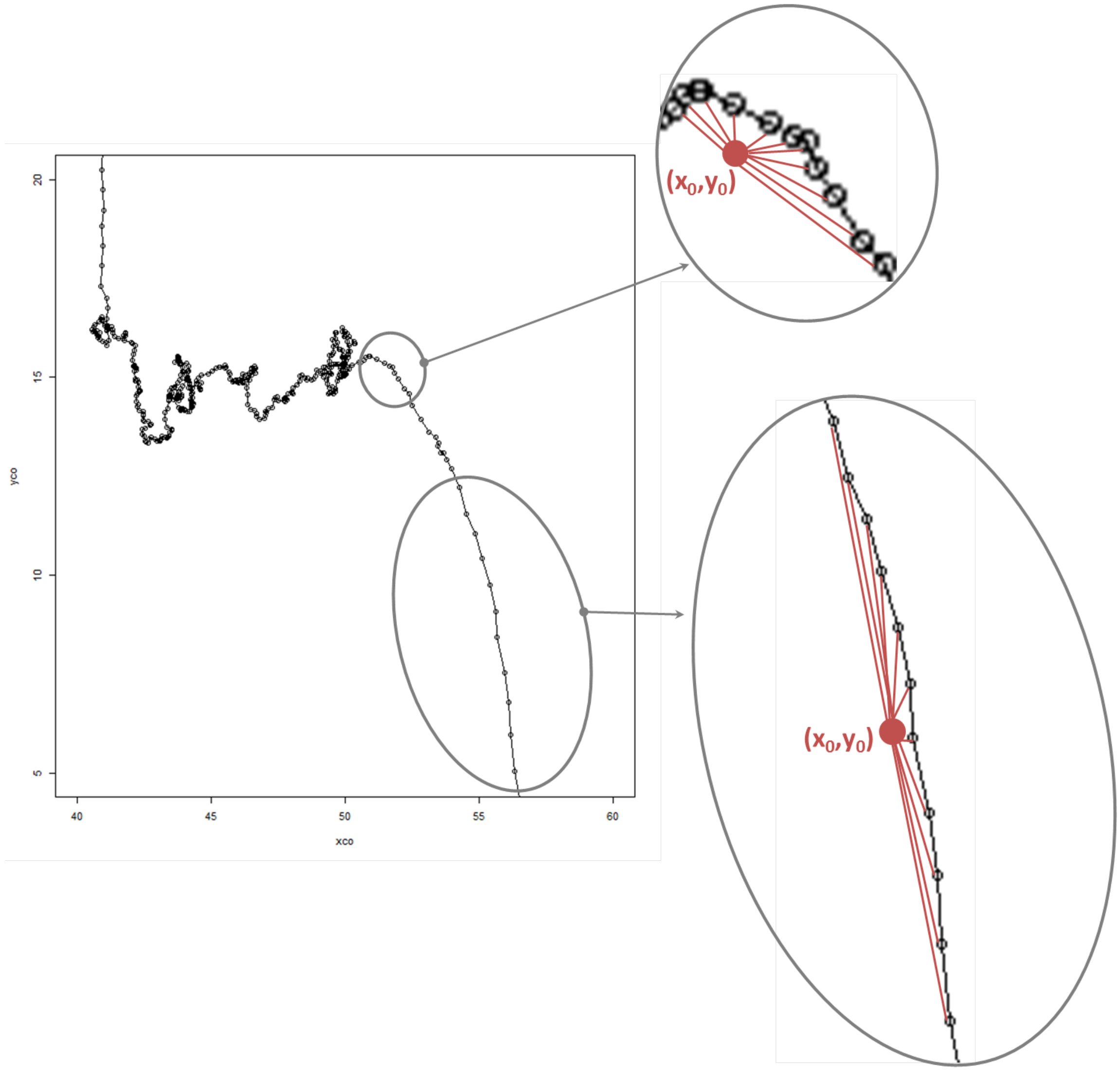


Figure 5: An example of two windows of 11 position points being considered whilst the system is out-of-control. The mean position from these n points is found and taken to form the origin (x_0, y_0). Distances of each of these 11 position points from (x_0, y_0) are measured and if enough of these successive distances are suitably close together, the system is defined to be in-control.

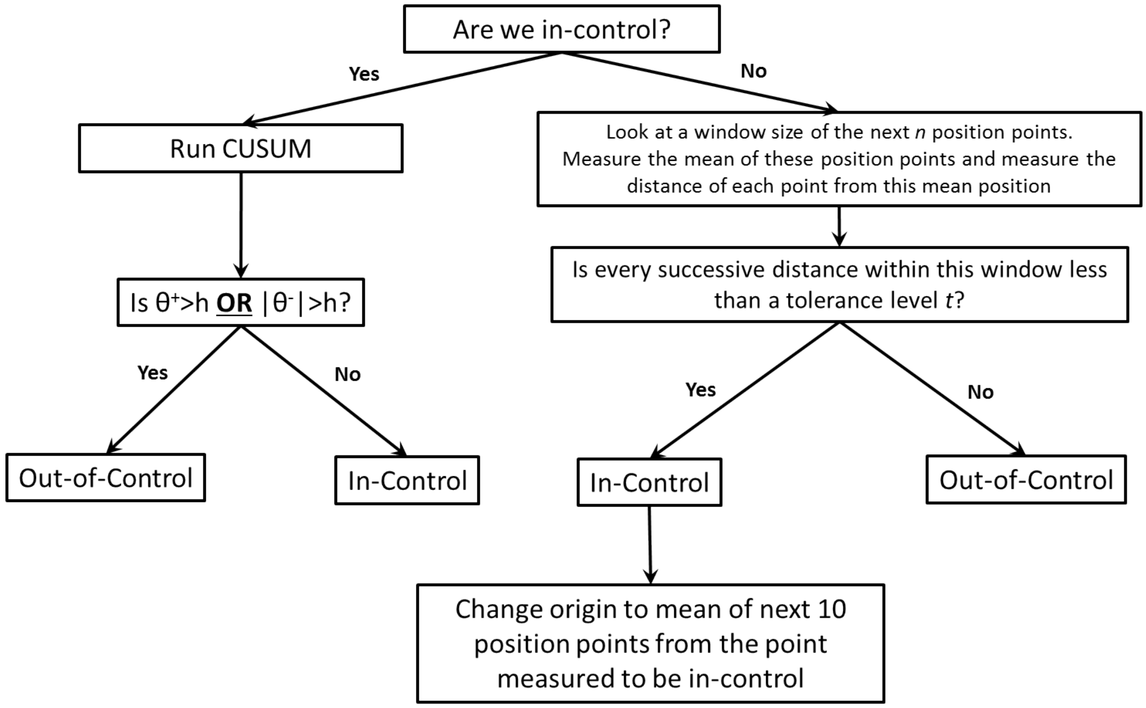


Figure 6: A decision tree outlining the algorithm used to monitor the state of the system through time

4 Developing the Hybrid CUSUM method for applications with animal movement data

4.1 The Data Set

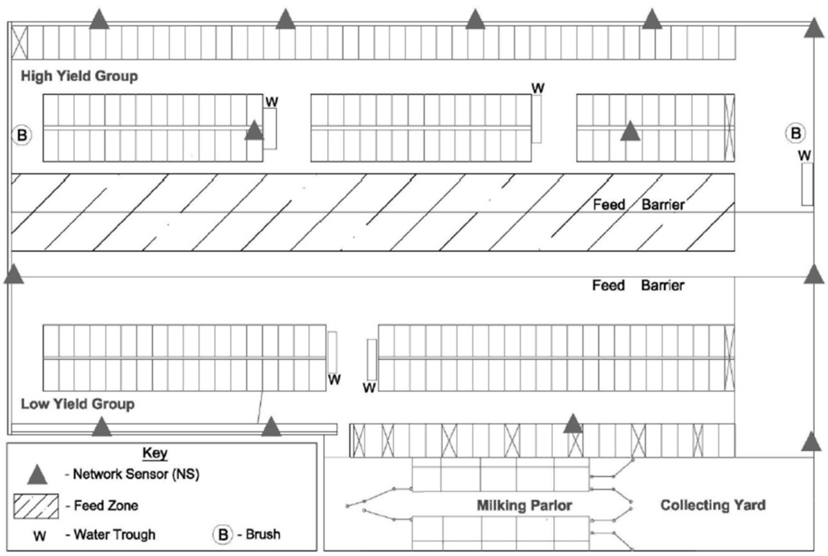


Figure 7: A map of the barn. The cows considered in this report had the resting area at the top of the map, accessed the feeding area across the center of the map and were taken off to be milked down a passage on the right-hand-side of the map to a separate milking area in the bottom-right of the map.

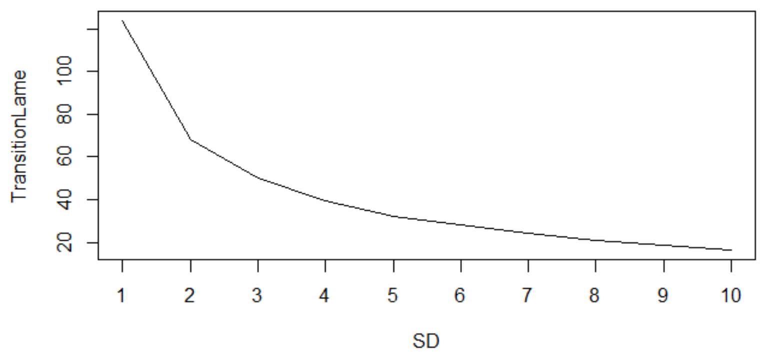
The Hybrid CUSUM outlined in this report has been applied to a data set of Holstein dairy cows housed within a commercial farm in Essex, UK. The data set was collected as part of a collaboration between Writtle University and the University of Essex and consists of 10 lame cows and 10 non-lame cows which were selected from a much larger herd and tracked for seven days from the 22^{nd} to the 28^{th} of January. Although data was collected for the full seven days, biologists monitoring the cows suggested that the data from the 22nd of January would not provide accurate results since the trackers had just been fitted and so the cows may not exhibit their true behaviour whilst they adjusted to the sensation of having a tracker attached to them. Data for the 28th of January was also deemed unusable by the biologists since the breed classifier was working at the farm on this day, resulting in the cows being out of the barn for long periods of time and thus disrupting their daily routine. As such, only data from the 23rd to the 27th of January was used on the advice of the biologists from Writtle University overseeing the collection of the data. Upon closer inspection of the data, it was found that sensor information for one non-lame cow was not valid for four out of the five days of collection and so the decision was made to remove this cow from the analysis, leaving ten lame and nine non-lame cows left in the study.

Data was collected every eight seconds using an Omnisense Series 500 Cluster Geolocation System (Omnisense, 2013) which enabled relative position and tri-axial accelerometer signals to be collected wirelessly through a network of mobile sensors attached via collars to the cows' necks and stationary sensors fixed to points in the barn. The data was converted into CSV format and produced a full time and date code to a millisecond level of accuracy, the unique identification code of the sensor, three-dimensional position coordinates relative to a corner of the barn and the mean and peak accelerometer signal collected over the eight seconds. Alongside this, observation data from biologists monitoring the cows' movements, milk yields and medical records were also provided.

The layout of the barn consisted of three zones (Figure 7): a resting area with open cubicles; an open feeding passage spanning the width of the barn; and a separate milking area. Although Figure 7 shows cubicles and feeding area access from both sides of the barn, the cows considered in this report were housed in the top half of the barn. The daily routine of the cows involved them being separated into groups to be milked three times a day, broadly corresponding to morning, noon and evening.

4.1.1 Standard deviation

A rudimentary value of the standard deviation can be set depending on the size of the animal being considered and the amount of noise and error in the system. In the application considered in this report, dairy cows are of interest which are typically around two metres making a standard deviation of 2 a logical value to initially be used, although further exploration concludes that a standard deviation of 2 is slightly too small (see Figures 8 and 9). The data has also been smoothed (Section 4.2.2) which means that there is not as much noise in the system for the standard deviation to take account for, putting more emphasis on the size of the animal being considered for the value of the standard deviation. As a form of verification of the rudimentary value for standard deviation and to provide more certainty, the CUSUM can be run with varying values of standard deviation and the number of residencies and transitions detected can be plotted (Figure 8). A justification for the range of values for this parameter could be used such as that seen in Knell & Codling (2012) which argues that a standard deviation in the region of 2 to 4 would be appropri-



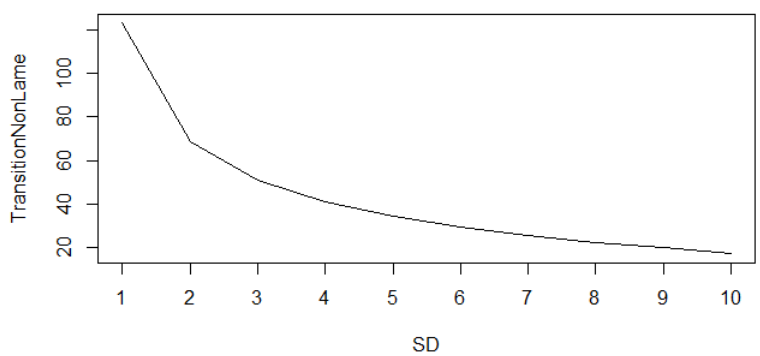
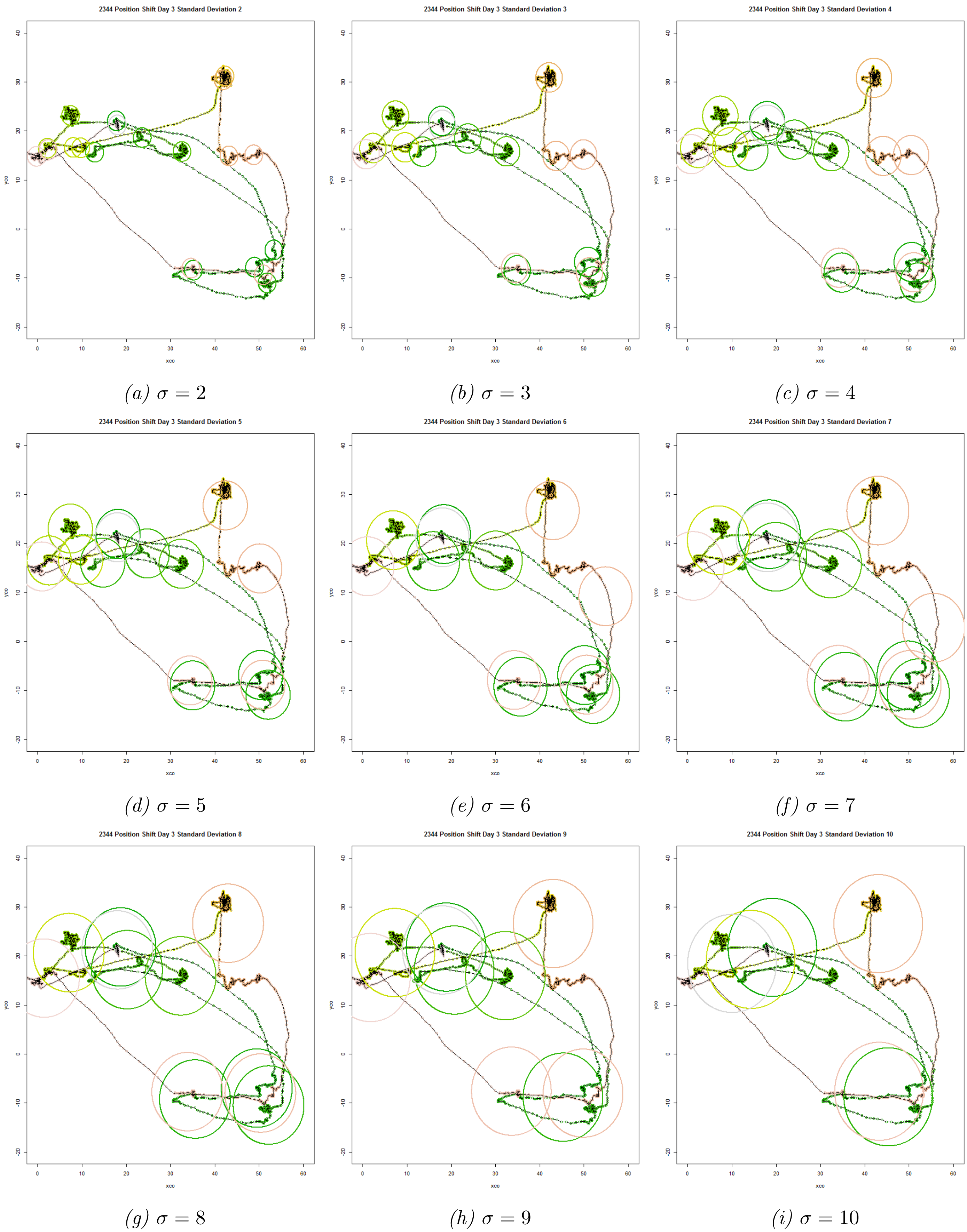


Figure 8: The amount of transitions against values of standard deviation (top) and the amount of transitions against values of standard deviation (bottom)

ate since values below this are detecting high numbers of residencies and transitions that are likely noise or small deviations by the animal, whilst standard deviation values above this detect much fewer residencies and so are missing key small transitions.

The technique used to verify the standard deviation in this report however, was to use the two-dimensional position data over the course of a day. The CUSUM results could be used to identify the location of each of the residencies and circles with radii defined by the standard deviation centered around each detected residency location and superimposed over the two-dimensional position plot (Figure 9). The result was a clear visual verification of whether the residencies were big enough to contain typical slight deviations of the animal whilst in a residency or too large that transitions were not being included. The conclusion was that a standard deviation of 3 should be used for this particular application.

Notice that for larger standard deviations, some residencies are not centered exactly on the obvious residency location. This is because a larger standard deviation value allows more deviation between successive distances in order to be defined as in a residency state from a transition state which means that the system is defined as being in-control earlier. A larger standard deviation also allows more movement within a residency state which means that it takes longer for the animal to be detected as entering a transition state, thus it takes longer for the system to be detected as entering an out-of-control state. Both of these situations will cause the center of the residency to be slightly off-center of the “true” residency location.



(a) $\sigma = 2$ (b) $\sigma = 3$ (c) $\sigma = 4$

(d) $\sigma = 5$ (e) $\sigma = 6$ (f) $\sigma = 7$

(g) $\sigma = 8$ (h) $\sigma = 9$ (i) $\sigma = 10$

Figure 9: 2D position data with circles centered around the location of each detected residency. The radius of each circle is given by the standard deviation and gives a visual indication of the allowable deviation of the animal before being classified as a transition state. The circles have been coloured to match the colour coding of where the animal is situated through time.

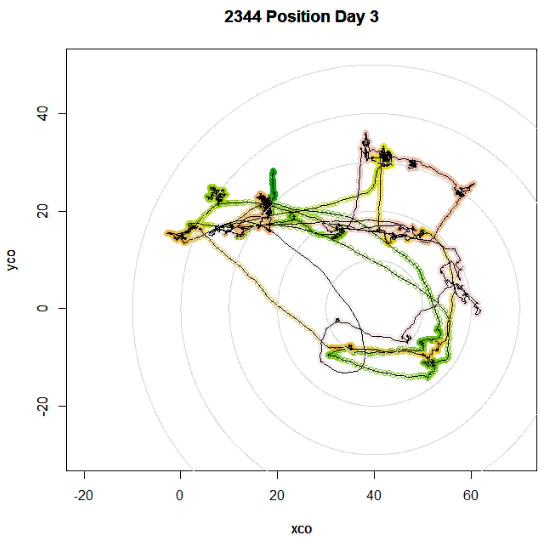
4.2 Initial Data Analysis and Processing

4.2.1 Position and Accelerometer Data

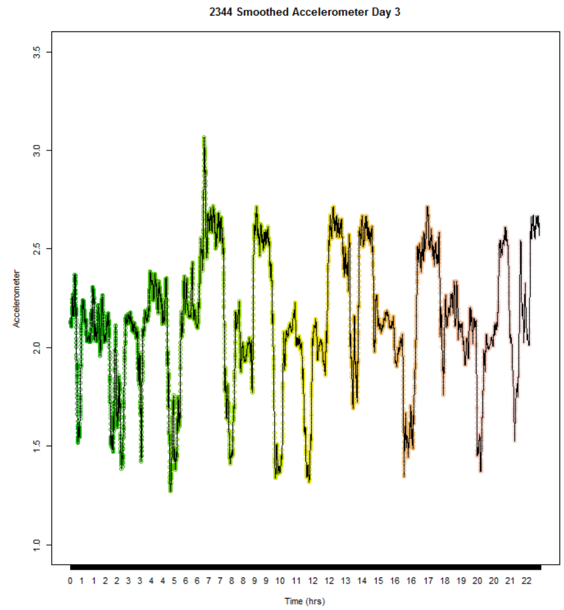
The data used in this report consists of twenty dairy cows with accelerometer and three-dimensional position data being collected every eight seconds over the course of seven days (Section 4.1). Initial data analysis shows the plots of the x and y coordinates of each cow over the course of a day to show its position. This can be further improved by applying a colour gradient to the points according to time to further illustrate the cow's path over the course of the day (Figure 10a). Although not the focus of this report, plotting the accelerometer signal against time can give an indication of activity over the course of the day with low activity broadly corresponding to resting and high activity broadly corresponding to feeding (Figure 10b).

It can be seen in Figure 10a that there are clusters of points where a cow is staying in the same location for prolonged periods of time. These are the residency periods which the Hybrid CUSUM is designed to accurately detect. It is also interesting to note that the lower activity levels in Figure 10b correspond to the same residency periods in Figures 10a and 10c; the amount of activity during these periods determines the type of activity (resting, feeding, milking) that the cow is doing. It is also important to note that a residency in the feeding area will probably have higher activity than a residency in the resting area.

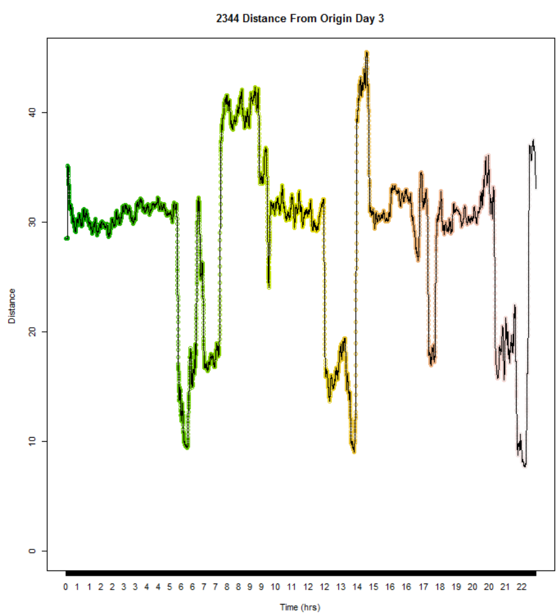
One apparent difficulty to overcome when working with the data is that sensor changes would frequently occur meaning that a cow may start the day with one sensor but finish the day with another sensor. These sensor changes were common due to either a signal loss (Section 4.1) or simply because the battery had run out of charge. Whilst sensor information was monitored as closely as possible for just such situations and sensors changed as promptly as possible, there are instances where the data is not valid or days where the cow information is spread across two sensors. Since some analysis has already been carried out with the data and the data has had to be cleaned before use, it was deemed sufficient to remove data which had already been deemed unusable for other studies without the need for extra first-hand cleaning. As an example, one non-lame cow was completely removed from further analysis since four of the five days resulted in a signal loss.



(a) *x and y position data*



(b) *Accelerometer data*



(c) *Distance from the origin point (40,0)*

Figure 10: An example of the data collected by the sensors for one cow over the course of one day. Figure 10a shows plot of x and y position data, Figure 10b shows a plot of accelerometer data and Figure 10c shows a plot of the distance from an arbitrary fixed origin position (40,0). Figures 10a, 10b and 10c have been smoothed using a 30 time point moving window.

4.2.2 Smoothing

Although this data set is of very high resolution with data being collected every eight seconds, there is also a lot of noise within the data due to the construction of the barn and the way that the data was collected. Location data was collected by measuring the relative distance of each of the mobile sensors from each other with the exact positions being pinpointed by the fixed stationary sensors. The barn itself consists of a multitude of metal structures such as the cubicles and the feeding area which resulted in signals from the sensors being reflected off of the metal surfaces creating noise. As such, a suitable smoothing window is required to remove this level of noise from the data.

The smoothing applied to the data was by using a moving window of a fixed sized n that takes the average of all x ordinates and y ordinates within this window to produce a smoothed coordinate at point i:

$$(x_i, y_i) = \left(\frac{\sum_{k=i}^{i+n} x_k}{n}, \frac{\sum_{k=i}^{i+n} y_k}{n} \right)$$

The size of the window to be considered was varied between 5 and 100 time points to see which would give data which is smoothed enough to remove unwanted noise, whilst not smoothed too much to remove important details. A window size of 30 time steps was deemed appropriate which corresponds to 4 minutes worth of smoothing. Whilst there is still some noise present with this smoothing window, enough fine trajectory detail is left and a CUSUM process should act further to smooth some of the noise in the system.

As a further verification, different smoothing windows were run through the Hybrid CUSUM algorithm and the number of transitions and residencies detected was plotted (Figure 11). For smoothing windows between 0 and 20 time points, large amounts of noise is still present in the data and so the algorithm is detecting the system to regularly switch between in-control and out-of-control states. For smoothing windows above 40 time points, less noise is present

4.3 Testing the back in-control window size

When the Hybrid CUSUM detects the system to enter a state of out-of-control, a second algorithm is initiated to detect when the system re-enters a state of in-control again (Section 3.5.2). The original algorithm for doing this can be found in Appendix A and was not retained since it falsely detected residency periods which were verified by using a similar plot to Figure 9. The improved algorithm used in this report identifies when the system is back in-control by using a window size of n points and taking the mean of these n coordinates to be the origin (x_0, y_0). The distances of these n points from this origin position is calculated and if m of these successive distances are within a given tolerance level t, then the system is defined to be in-control. Otherwise the algorithm is advanced to the next time point.

The three parameters that needed to be considered here are the number of points in the window n, the number of these points needed to be suitably close together m

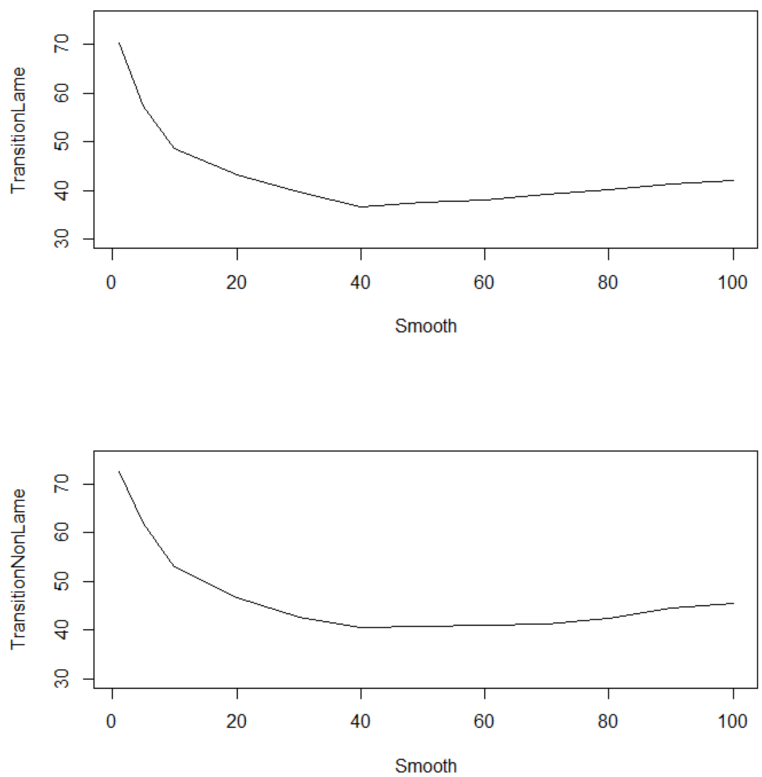


Figure 11: The total number of transitions tested against varying smoothing window sizes for lame cows (top) and non-lame cows (bottom)

and the acceptable tolerance level which these successive points need to be within t. The method for choosing the value of the standard deviation was to use circles with the same radius as the standard deviation, centered on the residency position, overlaid on the position trajectory plot (Figure 9). This provided a visual verification that could be used to adjust the standard deviation to include the cow's expected deviation within a residency state. A similar method could be used to verify the parameters used here and to prevent large amounts of false-positive residencies from being detected.

It was quickly realised that the optimisation process could be reduced by imposing the condition that every point within the window of size n had to be within the tolerance level which meant that the parameter m was equal to n. From this, the parameter t could be justified. Since differences between the successive distances were being standardised, a t of 1 would correspond to an allowance that successive points could be within 3 metres of each other; each successive time point is 8 seconds apart so a t value of 1 would mean that the cow has to travel less than 3 metres in 8 seconds in order to be defined as in-control. Given that the aim is to detect when the cow is starting to slow down, the successive distances should be suitably smaller than the distance which a cow is expected to travel in 8 seconds. Given also that a dairy cow does not move relatively quickly, it is perfectly possible that a cow may indeed only travel 3 metres in 8 seconds, as such a t value less than 1 is required. It was concluded that a t value of 0.1 which corresponds to 0.3 metres was appropriate since if a cow had entered a state of residency and so successive distances should only deviate by a small amount. Certainly, if a cow is in a transition state, it should travel more than 0.3 metres in 8 seconds.

Whilst changing the value of t, the number of points that were being considered n was also changed. Since all points within this window were required to be within the tolerance t, then smaller values of n will detect short residencies which may be useful in the context of this report when analysing the difference in behaviour between lame cows and non-lame cows but this has the added disadvantage that some of these short residencies would be noise. Alternatively, large values of n will detect longer residencies and will likely remove large amounts of noise, however short residencies will not be detected. As such, two values of n were chosen: 11 time points which corresponded to 88 seconds (approximately one-and-a-half minutes) and would detect the shorter residencies; and 75 time points which corresponds to 600 seconds (10 minutes) and would detect the longer residencies.

Once the system is defined to be in-control, the algorithm states that the position in middle of the window is the most appropriate point at which the system is back in control, since it is the point closest to the mean position of the window (x_0, y_0) which the cow was actually at.

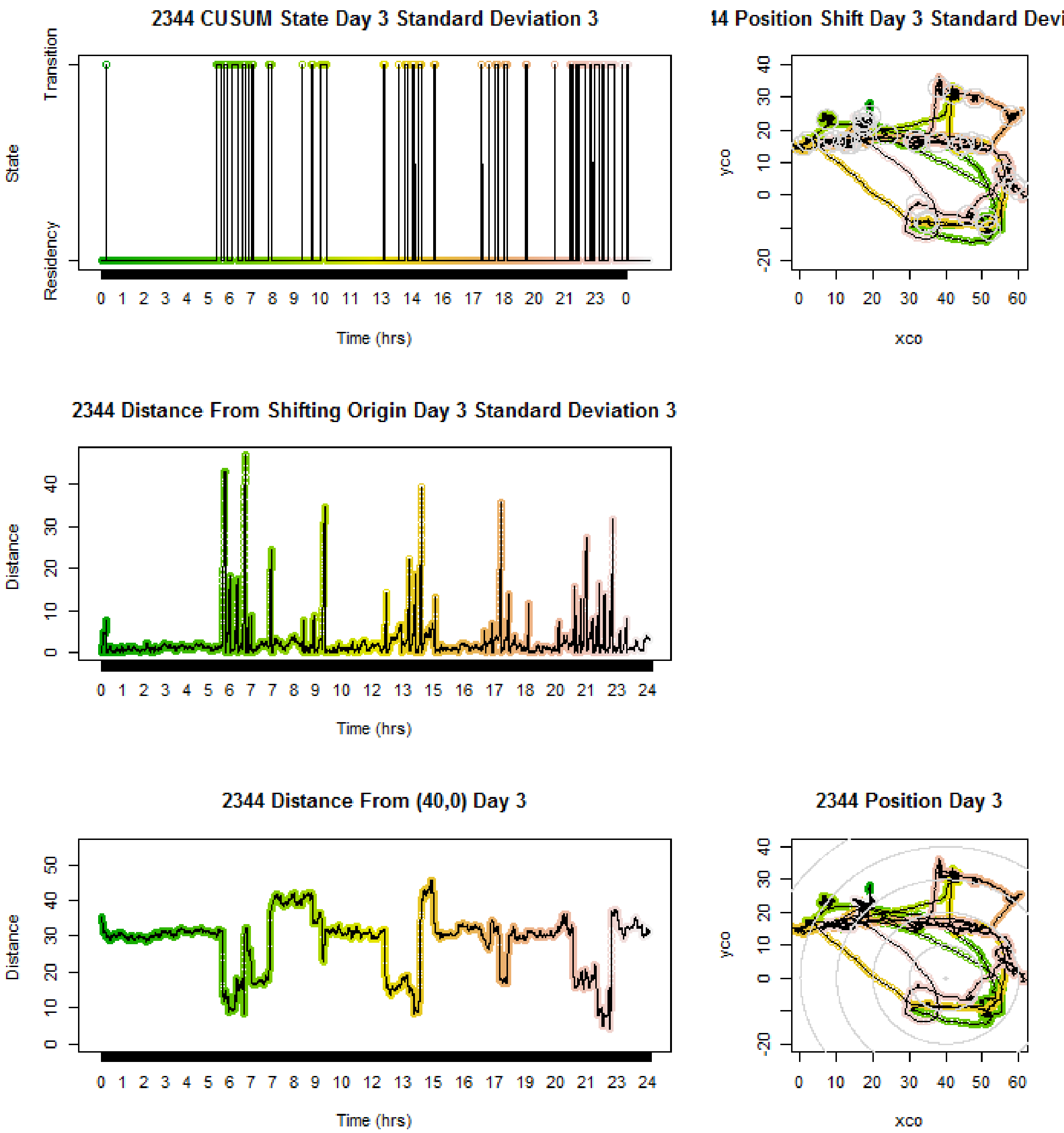
4.4 Summary Data Collected from the Hybrid CUSUM Algorithm

The data used in the Hybrid CUSUM was x and y position data correct to two decimal places (0.01 meters) and time and date code correct to the nearest second (recognising that data was collected every eight seconds). This allowed the CUSUM to collect data about where the cow was in a state of residency in the barn but also when the cow was in a state of residency or transition. From this, information such as the duration of a residency or transition and the types of the transitions (such as moving from the feeding area to the resting area) can be inferred.

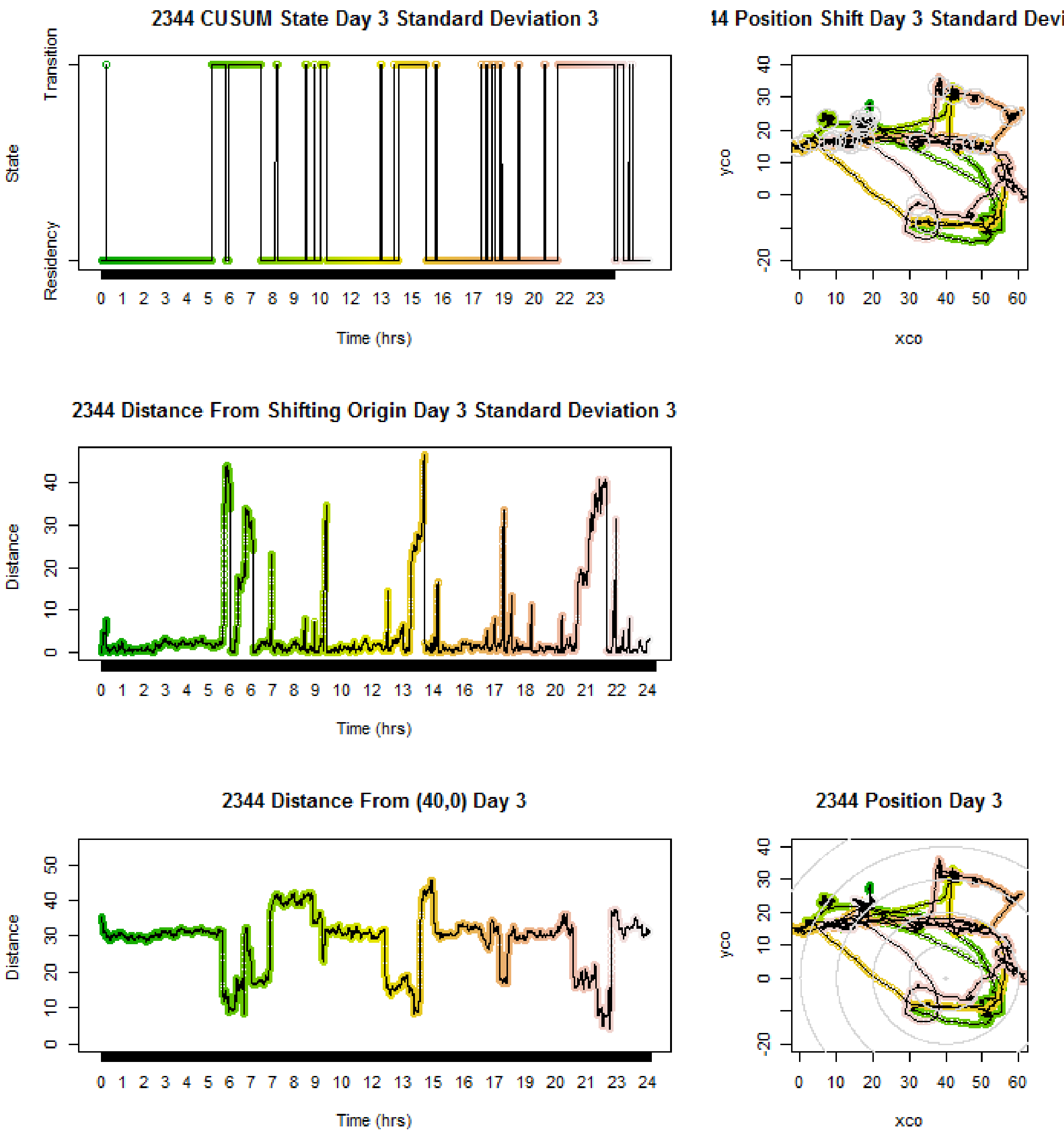
From the time point when the CUSUM detects the system to be back in-control, the time code is recorded to be the start of a residency and the end of a transition. The CUSUM then runs until it detects a state of out-of-control, at which point the running mean used whilst the system was in-control is recorded as the residency position for that period and the time code is recorded to be the end of a residency and the start of a transition. The duration of the residency can be calculated as being the amount of time between the start of a residency and the end of a residency, likewise for transition duration to be the amount of time between the start of a transition and the end of a transition.

Identifying whether each residency is in the resting area, feeding area or milking area can be achieved by looking at the map of the barn (Figure 7) and identifying that a y-ordinate: above 20 should broadly correspond to the resting area; between 10 and 20 should broadly correspond to the feeding area; and below 10 should broadly correspond to the milking passage or milking parlour (which for convenience will be classified as the milking area). The residency positions are then classified using these criteria to give each residency as either in the resting area, feeding area or milking area. Transitions between each zone can be inferred by considering consecutive residency states since a transition must have occurred between each residency. It is important to note that this method is only able to identify that a cow was resident within the resting area, feeding area or milking area rather than being able to specify that it was explicitly resting, feeding or being milked since it is only location data that is being used. For example, a cow may be resident in the feeding area but not doing the activity of feeding. Combining the CUSUM approach with an accelerometer signal approach such as that seen in Diosdado et al. (2015) will be able to identify with more certainty that a cow is indeed resting, feeding or milking in each of the separate zones. It was not deemed necessary to use such analysis since the scope of this report is to develop a method to identify residency and transition behaviour based solely on position data.

In order to provide a visual way of displaying the result from the CUSUM of whether a cow is in a state of transition or residency through time, a **0** is recorded for every time point when the system is defined as being in-control (i.e. a state of residency) and a **1** is recorded for every time point when the system is defined as being out-of-control (i.e. a transition). The value of this counter is then plotted against time to give a binary plot for all the time points when the cow is in a residency and a transition. This visual verification is enhanced when directly compared to the distance from a fixed origin in the barn and the distance from the most recent residency location since it is possible to see that periods which are horizontal or flat on the distance plots imply that a cow was not moving and so should be plotted as **0** on the binary CUSUM plot, whereas the periods which are spiked and show clear movement on the distance plots should be plotted as **1** on the CUSUM plot (Figure 12).



(a) A CUSUM plot using an out-of-control window size of 11 time points



(b) A CUSUM plot using an out-of-control window size of 75 time points

Figure 12: Plots showing the binary nature of a CUSUM: ***0*** *for an in-control/residency period and* ***1*** *for an out-of-control/transition period. Notice that with 11 time points the CUSUM detects smaller residencies and so the CUSUM plot switches more frequently between in-control and out-of-control whereas with 75 time points only the larger residencies are detected and as such the CUSUM remains in a state for longer.*

5 Results

5.1 Hybrid CUSUM consistency with observation data

5.1.1 Method to check consistency

Whilst the data was being collected, a team of biologists also collected observation data for a portion of four of the five usable days which can be used to verify the consistency of the Hybrid CUSUM. The observation data was collected by using a two-dimensional grid overlaying a map of the barn and then each of the twenty cows were plotted in the square that they were observed to be closest to. The activity of the cow was also noted, although this type of observation data is more directed to the accelerometer data and so is of limited use for the application of the CUSUM described here. This observation took place over successive twenty-minute windows and it is important to note that only the start time of each window was noted meaning that the observation could have actually been made at any time during a twenty-minute window.

This data has been used to monitor the consistency of the CUSUM results with the observed data collected by the biologists. Using this observation data, it is possible to extract whether a cow has moved positions between two successive twenty-minute windows, thus implying that a transition has taken place, otherwise suggesting that the cow has remained in a state of residency. Note that the observation data only *suggests* whether a residency has taken place since it is feasible for a cow to have left the recorded position during an observation window, thus entering a transition state, but returned to the same position before the next observation window. In such situations, the observation data would suggest a residency to have taken place, when in reality the cow had been in a transition state. This requires more emphasis to be placed on the transitions extracted from the observation data rather than the residencies.

Extracting a residency or transition state from the observation data is achieved by taking the observed position coordinates of two successive observation windows $t_1 = (x_1, y_1)$ and $t_2 = (x_2, y_2)$ and measuring the change in distance D from each other:

$$D_{t_1,t_2} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

If this distance D is less than or equal to a given tolerance then it is concluded that the observation data has recorded no transition occurring, whereas if D is greater than the given tolerance then the observation data has shown a transition occurring. A sensible value for the tolerance level is 3 since this corresponds to 3 metres which is the same as the chosen standard deviation of the Hybrid CUSUM.

Since the data collected from the Hybrid CUSUM includes the start and finish times of transition and residency periods, it is possible to assign every time of the day to either a transition or residency state for each cow. If the observation data has suggested a transition to have occurred between two successive twenty-minute windows, this allows a forty-minute window for the cow to have entered a transition state. The times of this forty-minute window are then compared with the transition times recorded by the Hybrid CUSUM for the same cow to check for consistency. For convenience, a counter system is used to count every transition suggested by the observation data with a **1** and then assigning a separate counter with either a **1** if the Hybrid CUSUM has also identified a transition or a **0** otherwise. The total number of **1**s recorded by each counter are compared to give a percentage of transitions identified by the CUSUM when the observation data also identifies a transition.

The same consistency check was repeated with the residency states identified by the observation data and comparing whether the Hybrid CUSUM has also detected a residency state during this period. Care needs to be taken with interpreting the results of this approach however, since the observation data was recorded by approximating the location of the cow to the nearest square on the grid which was overlying the plan of the barn. It may have been, for example, that the resolution of the grid was not overly high and so the cow may have been in a transition state for the forty-minute period (two-successive twenty-minute windows) but the entire transition took place within the same square of the grid used to overlay the map of the barn and so was suggested as being resident by the observation data. The other alternative was that the cow was recorded as being in a location in one observation window then moved to another part of the barn but came back to the same location to be recorded as not having moved for the next observation window. As such, both of these scenarios would have recorded the cow as resident for the forty-minute period by the observation data, when in fact the CUSUM data would have detected it as being in a state of transition. This would not happen with the transition consistency check since if the location of a cow is recorded to be different between two successive observation windows, then the cow must have moved and so must have entered a transition state. This means that the residency consistency check may not give as consistent results as the transition consistency check.

5.1.2 Results of the consistency check compared with the Hybrid CUSUM

	Observed Residencies	**Detected Residencies**	**Consistency Percentage**	**Observed Transitions**	**Detected Transitions**	**Consistency Percentage**
Cow 2344 Day 2	7	0	0	9	9	100
Cow 2010 Day 2	5	3	60	11	11	100
Cow 2060 Day 2	8	3	37.5	8	8	100
Cow 1891 Day 2	5	1	20	12	12	100
Cow 1078 Day 2	8	2	25	8	8	100
Cow 2616 Day 2	9	1	11.11	7	7	100
Cow 2302 Day 2	11	2	18.18	5	5	100
Cow 2003 Day 2	8	4	50	8	8	100
Cow 1184 Day 2	9	0	0	7	7	100
Cow 1340 Day 2	6	3	50	10	10	100
Cow 2959 Day 2	6	1	16.67	10	10	100
Cow 2512 Day 2	6	4	66.67	9	9	100
Cow 2472 Day 2	5	1	20	11	11	100
Cow 2179 Day 2	5	1	20	11	11	100
Cow 2172 Day 2	6	3	50	10	10	100
Cow 2153 Day 2	8	3	37.5	8	8	100
Cow 1892 Day 2	8	1	12.5	7	7	100
Cow 1491 Day 2	4	0	0	12	12	100
Cow 2596 Day 2	8	2	25	8	8	100

	Observed Residencies	Detected Residencies	Consistency Percentage	Observed Transitions	Detected Transitions	Consistency Percentage
Cow 2344 Day 3	0	0	NA	11	11	100
Cow 2010 Day 3	0	0	NA	11	11	100
Cow 2060 Day 3	1	0	0	10	10	100
Cow 1891 Day 3	0	0	NA	11	11	100
Cow 1078 Day 3	1	1	100	10	10	100
Cow 2616 Day 3	0	0	NA	11	11	100
Cow 2302 Day 3	1	1	100	10	10	100
Cow 2003 Day 3	0	0	NA	11	11	100
Cow 1184 Day 3	1	1	100	10	10	100
Cow 1340 Day 3	0	0	NA	10	10	100
Cow 2959 Day 3	0	0	NA	11	11	100
Cow 2512 Day 3	1	0	0	10	10	100
Cow 2472 Day 3	0	0	NA	11	11	100
Cow 2179 Day 3	1	1	100	10	10	100
Cow 2172 Day 3	0	0	NA	11	11	100
Cow 2153 Day 3	1	1	100	9	9	100
Cow 1892 Day 3	0	0	NA	11	11	100
Cow 1491 Day 3	0	0	NA	11	11	100
Cow 2596 Day 3	1	1	100	10	10	100

	Observed Residencies	Detected Residencies	Consistency Percentage	Observed Transitions	Detected Transitions	Consistency Percentage
Cow 2344 Day 5	1	1	100	13	13	100
Cow 2010 Day 5	0	0	NA	14	14	100
Cow 2060 Day 5	1	0	0	13	13	100
Cow 1891 Day 5	1	0	0	12	12	100
Cow 1078 Day 5	2	1	50	12	12	100
Cow 2616 Day 5	1	0	0	12	12	100
Cow 2302 Day 5	0	0	NA	14	14	100
Cow 2003 Day 5	0	0	NA	14	14	100
Cow 1184 Day 5	2	1	50	12	12	100
Cow 1340 Day 5	2	2	100	12	12	100
Cow 2959 Day 5	0	0	NA	13	13	100
Cow 2512 Day 5	0	0	NA	14	14	100
Cow 2472 Day 5	1	1	100	13	13	100
Cow 2179 Day 5	0	0	NA	14	14	100
Cow 2172 Day 5	0	0	NA	14	14	100
Cow 2153 Day 5	1	1	100	13	13	100
Cow 1892 Day 5	0	0	NA	14	14	100
Cow 1491 Day 5	2	2	100	12	12	100
Cow 2596 Day 5	0	0	NA	13	13	100

	Observed Residencies	**Detected Residencies**	**Consistency Percentage**	**Observed Transitions**	**Detected Transitions**	**Consistency Percentage**
Cow 2344 Day 6	5	3	60	26	26	100
Cow 2010 Day 6	3	3	100	30	30	100
Cow 2060 Day 6	1	0	0	32	32	100
Cow 1891 Day 6	2	2	100	30	30	100
Cow 1078 Day 6	2	1	50	31	31	100
Cow 2616 Day 6	2	2	100	31	31	100
Cow 2302 Day 6	4	1	25	28	28	100
Cow 2003 Day 6	2	2	100	31	31	100
Cow 1184 Day 6	6	3	50	27	27	100
Cow 1340 Day 6	1	0	0	32	32	100
Cow 2959 Day 6	2	2	100	30	30	100
Cow 2512 Day 6	1	0	0	32	32	100
Cow 2472 Day 6	3	2	66.67	31	31	100
Cow 2179 Day 6	2	2	100	31	31	100
Cow 2172 Day 6	3	3	100	28	28	100
Cow 2153 Day 6	0	0	NA	32	32	100
Cow 1892 Day 6	4	2	50	30	30	100
Cow 1491 Day 6	2	0	0	31	31	100
Cow 2596 Day 6	1	1	100	32	32	100

	Observed Residencies	**Detected Residencies**	**Consistency Percentage**	**Observed Transitions**	**Detected Transitions**	**Consistency Percentage**
All Cows	200	79	39.5	1193	1193	100

Table 1: Comparing consistency of the observation data with the CUSUM data for an out-of-control window size of 11 time points.

The results for this consistency check can be seen in Table 1 where the number of transitions and residencies suggested by the observation data is given and then comparing whether the same state is identified by the Hybrid CUSUM for the same observation period to give a percentage. It can be seen that every transition period identified by the observation data was also detected by the Hybrid CUSUM which means that our CUSUM consistently identifies out-of-control periods. The residency periods have a much lower consistency but this is probably due to the data collection method of the observation data (as described above) rather than necessarily a failure of the Hybrid CUSUM. It is important to note that other verification methods such as that seen in Section 3.5.1 where the residency locations identified by the Hybrid CUSUM were superimposed over the position plot of the cow (Figure 9) which clearly showed that the CUSUM correctly identifies residency periods.

5.2 Statistical Analysis

The main aim of this report is to develop a Hybrid CUSUM methodology for the application of movement analysis in order to identify the residency and transition behaviour of domestic animals. Consequently, the data set used in this report is primarily intended to test the method of the Hybrid CUSUM and provide an indication of how this method could be used when applied to animal data sets. Despite this however, it is interesting to see whether such methodology is able to detect biological differences in movement behaviour between lame and non-lame cows within the same herd.

One of the statistical tests carried out was using the Welch t-Test (Welch, 1947) since the aim is to test if there is a difference between the movements of a sample of ten lame cows and nine non-lame cows with null and alternative hypotheses:

H_0 : there is no significant difference between the means of the two samples

H_1 : there is a significant difference between the means of the two samples

A Welch t-Test was initially deemed to be appropriate for this application since the sizes of the two samples are relatively small and the two samples cannot be assumed to have equal variances. The Welch t-Test has to assume that the samples are normally distributed, and so the Shapiro-Wilks Test for normality was used (Shapiro & Wilk, 1965) with null and alternative hypotheses:

$$
\begin{aligned}
H_0 &: \quad \text{the sample is normally distributed} \\
H_1 &: \quad \text{the sample is not normally distributed}
\end{aligned}
$$

The outcome of the Shapiro-Wilks tests resulted in a small p-value for both samples at the 5% significance level which suggested that the majority of the data was not normally distributed and so the conclusion was made to substitute the Welch t-Test for the Wilcoxon Signed Rank Test which is a non-parametric test that does not assume the same normality condition as the Welch t-Test. The null and alternative hypotheses for the Wilcoxon Signed Rank Test are the same as the Welch t-Test:

$$
\begin{aligned}
H_0 &: \quad \text{there is no significant difference between the means of the two samples} \\
H_1 &: \quad \text{there is a significant difference between the means of the two samples}
\end{aligned}
$$

The results for the Welch t-Test and the Wilcoxon Signed Rank Test (Wilcoxon, 1945) gave very similar p-values and suggested significant differences in the same areas which provides evidence for the confidence of the results (the results for the Welch t-Test can be found in Appendix B for reference).

Since multiple Wilcoxon Signed Rank Tests are being carried out on the same data, the chance of getting a significant result being a false-positive (Type I Error) is increased so one must carry out a post-hoc correction procedure to suggest which significant results may be false-positives. The correction procedure used was the Benjamini-Hochberg (Benjamini & Hochberg, 1995) which is appropriate for this situation when multiple test have been carried out on the same sets of data. The Benjamini-Hochberg procedure ranks (r) the p-values from the Wilcoxon Signed Rank Test from smallest to largest and compares it to the total number of tests considered (n) and a chosen false discovery rate (Q) to produce a critical value:

$$\text{Benjamini-Hochberg Critical Value} = \frac{r}{n} \times Q$$

Since the Wilcoxon Signed Rank Test was carried out at the 5% level of significance, the false discovery rate of 5% was also chosen for the Benjamini-Hochberg correction procedure. The largest ranked significant p-value from the Wilcoxon Signed Rank Test which is less than the Benjamini-Hochberg critical value and all p-values ranked below this are accepted as being significant whilst all significant p-values above this are rejected as being false-positives.

5.2.1 Results of the Wilcoxon Signed Rank Test

Data from all of the cows was separated into two groups corresponding to the lame and non-lame cows and Wilcoxon Signed Rank Tests were used to identify if certain behaviours exhibited a difference. The decision was made to exclude a non-lame cow from the analysis since a large amount of data collected from this cow was deemed unusable due to sensor failure, leaving a sample of ten lame cows and nine non-lame cows. Multiple sensor changes took place over the course of the trial, mostly due to batteries needing charging. Whilst the approximate times of these sensor changes were recorded by the biologists, a period of twenty minutes before and after the recorded time of the sensor changes were removed to allow for discrepancies in the recorded times and to also allow time whilst the sensors recalibrated to the system after they were turned on. In addition to this, fifteen minutes of data was removed at the start of every day for all cows to account for the systems resetting and recalibrating. In a small number of cases, sensors failed for a number of hours resulting in either inaccurate data which was removed or a lack of data being recorded. In total, around 9% of data was either not recorded or removed for the analysis.

The tests used looked at summary data which the Hybrid CUSUM method could identify: total number of residencies and transitions; duration of residency periods in the resting, feeding and milking areas of the barn (Section 4.1); number of transitions between different areas of the barn. The same tests were carried out for data with an out-of-control window size (Section 3.5.2) of 11 time points and 75 time points to give the same analysis using shorter residencies and longer residencies respectively.

Out-of-control window size of 11 time points					
	W-value	**p-value**	**Lame Mean Value**	**Non-Lame Mean Value**	**Significant Difference in favour of**
Number of residencies	1257.5	0.9079	39.3	39.5	
Number of transitions	1257.5	0.9079	38.3	38.5	
Duration of resting residencies	1647	0.0001	5985.4	5315.9	Lame
Duration of feeding residencies	614	0.0001	2153.8	2553.1	Non-Lame
Duration of milking residencies	1154	0.8318	1121.2	1011.6	
Number of resting residencies	987.5	0.3061	14.0	15.7	
Number of feeding residencies	611.5	0.0001	14.9	17.8	Non-Lame
Number of milking residencies	1035.5	0.5036	10.4	10.0	
Resting to resting transitions	975	0.2635	7.1	8.5	
Resting to feeding transitions	775.5	0.0073	3.8	4.6	Non-Lame
Resting to milking transitions	1378	0.0459	2.2	1.8	Lame
Feeding to resting transitions	1080	0.7337	5.4	5.5	
Feeding to feeding transitions	692	0.0012	8.6	11.0	Non-Lame
Feeding to milking transitions	950.5	0.1630	0.8	1.0	
Milking to resting transitions	1053	0.5710	0.9	1.0	
Milking to feeding transitions	1290	0.1904	2.3	2.1	
Milking to milking transitions	1014	0.4059	7.1	6.9	

Table 2: Results of the Wilcoxon Signed Rank Test for an out-of-control window size of 11 time points carried out at the 5% level of significance.

Out-of-control window size of 75 time points					
	W-value	**p-value**	**Lame Mean Value**	**Non-Lame Mean Value**	**Significant Difference in favour of**
Number of residencies	1267	0.9591	21.7	21.7	
Number of transitions	1264	0.9428	20.8	20.9	
Duration of resting residencies	1592.5	0.0005	5646.3	4901.1	Lame
Duration of feeding residencies	715.5	0.0023	1651.8	2025.1	Non-Lame
Duration of milking residencies	1256.5	0.3288	570.8	451.5	
Number of resting residencies	1118	0.9611	9.8	9.5	
Number of feeding residencies	684	0.0009	8.3	10.0	Non-Lame
Number of milking residencies	1265.5	0.2863	3.5	3.3	
Resting to resting transitions	1152	0.8419	3.9	3.6	
Resting to feeding transitions	981	0.2700	3.3	3.7	
Resting to milking transitions	1178.5	0.6770	1.7	1.6	
Feeding to resting transitions	1196.5	0.5842	4.3	4.3	
Feeding to feeding transitions	661	0.0005	3.2	4.9	Non-Lame
Feeding to milking transitions	1100	0.8379	0.6	0.6	
Milking to resting transitions	1062	0.6214	0.9	1.0	
Milking to feeding transitions	1238.5	0.3774	1.6	1.4	
Milking to milking transitions	1197	0.5696	1.0	0.8	

Table 3: Results of the Wilcoxon Signed Rank Test for an out-of-control window size of 75 time points carried out at the 5% level of significance.

5.2.2 Results of the Benjamini-Hochberg Correction Procedure

The Benjamini-Hochberg correction procedure takes one of the ten p-values that were flagged as significant using the Wilcoxon Signed Rank Test and suggests it is a Type I error. The p-value in question is the number of resting to milking transitions for an out-of-control window size of 11 time points. It is known that most activity to and within the milking area is induced by farming routines rather than being voluntary behaviour by the cow and as such any significant differences between lame and non-lame cows in this area are unlikely to give an insight to behavioural differences between lame an non-lame cows. As a result, it can be justified with reasonable confidence that the Benjamini-Hochberg procedure has correctly identified the number of resting to milking transitions for an out-of-control window size of 11 time points as a Type I error.

The Benjamini-Hochberg procedure does support the conclusions of the Wilcoxon Signed Rank Test between the higher number of feeding residencies and transitions to and within the feeding area as well as duration of resting residencies so it can be concluded with relative confidence that non-lame cows tend to spend a larger proportion of time within the feeding area than lame cows and that lame cows typically spend longer time in the resting area.

	Out-of-control window size	p-value in Wilcoxon Signed Rank Test	Rank of p-value	Benjamini-Hochberg Critical Value	Significant Difference in favour of
Duration of resting residencies	11	0.0001	1	0.0015	Lame Cows
Number of feeding residencies	11	0.0001	2	0.0029	Non-Lame Cows
Duration of feeding residencies	11	0.0001	3	0.0044	Non-Lame Cows
Feeding to feeding transitions	75	0.0005	4	0.0059	Non-Lame Cows
Duration of resting residencies	75	0.0005	5	0.0074	Lame Cows
Number of feeding residencies	75	0.0009	6	0.0088	Non-Lame Cows
Feeding to feeding transitions	11	0.0012	7	0.0103	Non-Lame Cows
Duration of feeding residencies	75	0.0023	8	0.0118	Non-Lame Cows
Resting to feeding transitions	11	0.0073	9	0.0132	Non-Lame Cows
Resting to milking transitions	11	0.0459	10	0.0147	Lame

Table 4: Results of the Benjamini-Hochberg correction procedure carried out at the 5% false discovery rate.

6 Discussion

6.1 Hybrid CUSUM Method

The Hybrid CUSUM developed in this report has the ability to identify residency and transition periods from two-dimensional position data. Its consistency was verified with a 100% rate of detecting a transition when the observation data collected by the biologists also suggested a transition. When the same observation data was used to verify the consistency of residencies also detected by the CUSUM, its consistency was much lower at 39.5%. It was concluded that the observation data probably failed to correctly identify "true" residencies due to the possibility that cows could have left a location and come back to the same location between observation windows and so the observation data would suggest a residency, whereas the CUSUM would detect a transition. As such, an alternative method for verifying the detection of residencies by the CUSUM was used which consisted of superimposing the detected residencies over the position plot of a cow in the form of a circle with radius the size of the standard deviation and centre at the detected location of the residency. These circles represent the possible deviation that a cow can perform before entering a transition state and it was found that there were clearly circles containing all obvious periods where the cow was located in the same position for a prolonged period of time (Figure 9). This approach is similar to an Area Restricted Search method such as that seen in Kapota et al. (2017) where circles are used to overlay the animal's movement and to map the amount of time that an animal spends in a particular location in an aim to detect residency periods. The Hybrid CUSUM method in this report is more advanced in that it analyses an animal's distance to detect when an animal is resident or transitory and then uses these residency locations to center the circles.

Fauchald & Tveraa (2003) suggested a model to identify First Passage Time from an animal's movements by overlaying a circle roughly equal to a typical food patch and seeing how long that animal remains resident within it. This effectively determines large-scale residencies from animal movement, however the advantage of a Hybrid CUSUM compared to this method is that the location of such residencies can be identified with more accuracy. A Hybrid CUSUM also has the ability to identify shorter residencies and transitions which may be key to determining differences in animal behaviour, whereas the model suggested by Fauchald & Tveraa (2003) can only identify larger residency behaviour and may not identify differences in small-scale behaviour. The other difficulty is that there is a level uncertainty in the optimal circle size to choose with this method, whereas with a Hybrid CUSUM, the size of the deviation from a residency position can be justified using the biological length of the animal and from visual verification (Figure 9). Hidden Markov Models such as that seen in Patterson et al. (2008); Langrock et al. (2012) can also be seen as a useful way of determining residency and transition periods from a state-space model. The difficulty however is identifying the hidden state sequence (i.e. whether resident or transitory) from useful observable data which usually involves looking at the location of an animal in relation to its surroundings (e.g. in the context of the cow data, whether it is in the resting area, feeding area or milking area). By using a Hybrid CUSUM method, the only data that needs to be used is two-dimensional position data in order to extract information about the residency or transition state of the animal. The approach used in Knell & Codling (2012) was a partial sum approach based on a form of CUSUM to identify transitions between intensive and extensive search strategies of an animal. This method used animal movement to determine either speed or turning angle, whereas the Hybrid CUSUM in this report differs since its aim is solely to identify when an animal is stationary and when it is moving rather than show concern over *how* the animal is moving. It is possible to detect when an animal is slowing down and when it is speeding up by varying the out-of-control window size parameter (Section 3.5.2) since a small window size will detect periods in a transition when an animal is slowing down and record them as a residency. Whilst this is possible, the approach used in Knell & Codling (2012) is far more efficient to identify changes in behaviour when an animal is transitory.

Although CUSUMs have been used to detect an animal's activity such as that seen in Pastell et al. (2016) where the activity of a sow is measured using an accelerometer sensor in order to detect farrowing, the application of a CUSUM in this report uses position data as opposed to accelerometer data. The method in this report can also be used as a standalone method, whereas uses of CUSUMs with animal movement such as in Pastell et al. (2016) have had to rely on other models to be used in conjunction with a CUSUM. The two novelties of the Hybrid CUSUM method in this report are that the standard deviation is able to be predefined whilst the mean is calculated directly from the data and that the CUSUM is able to restart by imposing a second algorithm to identify when the system moves from a state of out-of-control to a state of in-control. This type of CUSUM is useful in the field of movement analysis to identify residency and transition periods of domestic animals from x and y position data. The algorithm has been enhanced further by imposing a shifting origin to measure the distance from the most-recent residency position as opposed to a fixed arbitrary origin. The fixed origin approach has the risk of situations where animals in a transition state are moving along contours and yet being measured as the same distance from the fixed origin, thus being detected as resident. The shifting origin reduces this risk, since even if an animal is moving along a contour it is the distance of the animal from the residency position that is of concern. As a result, the shifting origin can be applied more generally to other animal data sets.

Since this application looks at the ability to identify residency and transition periods, the standard deviation can be fixed to the expected amount of deviation that an animal makes whilst in a state of residency. Clearly a larger standard deviation will allow more deviation within the residency before being detected as a transition, however this may risk key small transitions not being detected. Alternatively a smaller standard deviation will identify key transitions but may falsely detect slight movement or noise within the residency as being in a transition state. The size of transitions may be key to identifying intensive foraging behaviour such as described in Knell & Codling (2012) so choosing a standard deviation that is small would be ideal. The method used to determine the correct size of standard deviation involved the initial justification that it should be a little larger than the animal being considered (i.e. around 2-3 metres for a cow). This was then verified further with a plot of the total number of residencies and transitions detected against changing standard deviation (Figure 8) and finding where the curve started to flatten out and then running the CUSUM with different values for the standard deviation and superimposing the detected residencies as circles with the radius of the standard deviation to see whether obvious residencies were completely contained within the circles (Figure 9).

With the standard DI- or SS-CUSUM, once a system is detected to be in a state of out-of-control the CUSUM switches off. With this Hybrid CUSUM, a second algorithm is initiated that looks at the clustering of a given number of successive position points and requires them to be within a certain tolerance of each other (Section 3.5.2). From the perspective of identifying residency and transition periods, this means that an animal is required to be stationary or close to stationary for the given number of points being considered before the system is defined as being back-in-control and for the CUSUM to restart.

Since all position points are converted from two-dimensional x and y coordinates into one-dimensional distance from origin (DFO) data, an origin needs to be defined. One solution is to fix an origin to always measure the distance from and was considered in Woor (2017) to be used with the cow data set considered in this report. This approach was fine for the cow data set since the cows were restricted by the layout of the barn to prevent them from moving along a contour (Figure 4a) and being recorded as a fixed distance (i.e. a state of residency) when in fact they are transitory, however the approach considered in this report used a shifting origin approach based on the most recent residency location and can be more-generally applied to other animal data sets where the animals are not restricted in their movements. This approach is deemed to be superior to the fixed-origin method since it can be applied more generally to animal data sets whose movements are not restricted by physical environmental barriers and the results of which are consistent both with the observation data taken by the biologists and also when superimposed over the position plot (Figure 9). Alternative applications of this Hybrid CUSUM approach to animal data may include identifying Home Ranges (Burt, 1943) or the position that an animal periodically revisits, however Home Range Analysis techniques such as that discussed in Section 2.2.4 are not directly relevant to the method used to detect residencies in this report since most Home Range Analysis techniques assume that an animal is already in a state of residency. By using a Hybrid CUSUM to identify residency periods and building a spatiotemporal plot of where an animal repeatedly remains resident over time such as that seen in Zhao & Jurdak (2016), it may be possible to determine routines from the positions of an animal's residency behaviour.

6.2 Computational Efficiency of Algorithm

The Hybrid CUSUM method may be constructed by primarily using a series of "if" statements to verify whether specific conditions are met (Section 4) along with relatively straightforward numeric calculations (Section 3.4) and as such is a relatively fast algorithm to compute. Whilst considering parameters such as computing power, choice of software and streamlining of the code may all contribute to improving the performance, it is certainly expected that computation should be on the order of minutes for a similar-sized data set to that seen in this report rather than days.

One aspect that may significantly improve computation is through the way that data is processed prior to the Hybrid CUSUM being applied. For instance, the cow data set used in this report had to account for sensor changes that occurred during the day which meant that data was located in separate files for each sensor and needed to be called in and appended to each other in order to be run as a full day of data. Since this was run within the same code as the Hybrid CUSUM algorithm then this slowed the overall computation time of the algorithm. A more-general processing technique that is likely to be common for most data sets is a smoothing method to reduce noise and, whilst this in itself is not necessarily overly computationally draining, will slow the computation time. An improvement which can be made then, is to consider whether simple processing may be performed within the sensor itself when data is being collected. Since the Hybrid CUSUM only requires the collection of x and y location data and a time stamp, it may be possible to direct some sensor computing power towards smoothing methods. Otherwise consideration towards processing the data in advance of running the Hybrid CUSUM would also act towards improving the efficiency and computation time.

Another important consideration is how well the method can be implemented to data sets with slower sampling rates (perhaps on the order of once a day) but collected over longer periods of time. In such cases, the behavioural characteristics such as the speed which the animal moves at and the area of a typical habitat needs to be known before it can be suggested whether this method would be suitable (Clark & Bjørnstad, 2004). For instance, if the sampling rate was once a day then clearly residencies which lasted for less time than this will not be detected, however if the animal migrates over large distances and remains in a location for weeks at a time then, providing a suitable value for standard deviation accounted for the typical habitat size, it may be possible that this algorithm would be suitable to detect when the animal is migrating and when it is resident.

6.3 Biological Implications

From the results in Section 5.2.1 there are significant p-values suggesting a difference in the feeding behaviour of dairy cows from a lame versus non-lame perspective. It is seen that by considering an out-of-control window size of 11 time points which corresponds to a requirement that the cow has to have stayed in relatively the same position for approximately one-and-a-half minutes before being considered as being in a state of residency, the non-lame cows spend significantly longer in the feeding area and also tend to transition a higher number of times within the feeding area which agrees with the findings in González et al. (2008); Norring et al. (2014) that lame cows spend less time feeding than non-lame cows. It was also found that lame cows spend a longer time in the resting area. From a biological perspective, it is thought that cows compete for better-quality food (Potter & Broom, 1987; Miller & Wood-Gush, 1991) and so the results of this report may suggest that non-lame cows move around the feeding passage more to find the better-quality food. A possible conjecture to be made from these results is that non-lame cows may behave dominantly and force the lame cows out of the feeding area, thus causing a significant difference between the duration of residencies within the feeding area for lame and non-lame cows.

The alternative out-of-control window size chosen consists of 75 time points which corresponds to a requirement that the cow has to have stayed in relatively the same position for approximately ten minutes before being considered as being in a state of residency, meaning that only larger residencies are being detected. By performing the same Wilcoxen Signed Rank Tests it is seen that the same significant differences are detected as with a window size of 11 time points, implying that the conclusions made about the results are consistent. Furthermore, many of the results also corroborate with the Welch t-Test (Appendix B) which, although not deemed statistically appropriate to be included within the main analysis of this report due to the failure of the data to be normally distributed using the Shapiro-Wilks Test (Section 5.2), still provides further confidence in the results if another significance test gives the same findings.

6.4 Further Exploration

This report has developed a Hybrid CUSUM method for use with animal data sets and applied it to a cow data set in order to identify periods of residencies and transitions and make comparisons between the movement behaviour of lame and non-lame cows. Whilst the conclusion of the report is that the method works well and there appears to be some interesting differences between the feeding behaviour of lame and non-lame cows, there is scope for some further exploration of the method. For instance, this report focused solely on utilising the two-dimensional position data of the cows which meant that when trying to identify the activity of the cow when in a state of residency, the assumption had to be made that the cow was doing the activity of feeding in the feeding area, resting in the resting area and milking in the milking area and whilst this is a perfectly reasonable assumption to make, its certainty may be improved by utilising the accelerometer signal and combining it with an approach seen in Diosdado et al. (2015) to verify what activity was taking place in each of the residency periods. Another possible extension of this is to use a similar CUSUM approach on the accelerometer signal to identify the residency states rather than the decision-tree algorithm utilised in Diosdado et al. (2015).

It may be interesting to develop the position of the residencies and the duration in each residency into a graph where the location of each residency is plotted onto a map of the barn in the form of a node. This could then be weighted so that the size of the node corresponds to the duration spent in that residency and arcs could connect the nodes to show the transitions between the residency locations. This could give a visual indicator of the differences in movement behaviour of lame cows versus non-lame cows. A similar method was performed in Zhao & Jurdak (2016) which looked at the spatiotemporal patterns performed by cattle using a two-state "stop and move" model. The "stop and move" model is very simple in comparison to the Hybrid CUSUM considered here in this report and so using the Hybrid CUSUM may give some more interesting insights.

This report used ten lame cows and nine non-lame cows and compared the data collected by the Hybrid CUSUM for these two groups in order to identify differences in the movement behaviour of lame and non-lame cows. This would be considered a herd-level perspective since it divides the herd into large two groups according to health status and directly makes comparisons as two individual groups. The most-natural extension would be to compare differences in data between individual cows as opposed to grouping multiple cows together. This may eliminate the risk of a single cow with dominant behaviour skewing results in favour of one particular characteristic; for instance, if one lame cow spends most of the days resident in the resting area, this may cause the entire herd of lame cows to be falsely detected as spending more time in the resting area when in fact this is just the particular dominant behaviour of one cow. Effect of individuals within dairy cows was carried out in Chua et al. (2002) where the welfare of calves kept in individual pens as opposed to with other calves was monitored to see the health benefits of each method. Although not directly related to this report, one of the key areas considered was movement which gives the potential for residency and transition behaviour and a Hybrid CUSUM could be used with accelerometer data. It does however highlight that individual-level analysis could give interesting insights. Other studies have shown that cows are not homogeneous (Howery et al., 1996) and so individual differences may be interesting to analyse as a further extension. The primary reason why this individual-level analysis was not performed in this report is that the scope of this report was to introduce a Hybrid CUSUM method to identify residency and transition periods with the application to an animal data set and then analysing whether the Hybrid CUSUM yielded any interesting differences between the behaviour of lame cows compared to non-lame cows. As such, the balance of the report is intended to offer both aspects in relatively equal proportions and so detailed individual-cow-level analysis was not deemed necessary.

References

Aebischer, N. J., Robertson, P. A., & Kenward, R. E. (1993). Compositional analysis of habitat use from animal radio-tracking data. *Ecology*, *74*(5), 1313–1325.

Akobeng, A. K. (2007). Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta paediatrica*, *96*(5), 644–647.

Baker, J. (2001). Population density and home range estimates for the eastern bristlebird at jervis bay, south-eastern australia. *Corella*, *25*(3), 62–67.

Barraquand, F., & Benhamou, S. (2008). Animal movements in heterogeneous landscapes: identifying profitable places and homogeneous movement bouts. *Ecology*, *89*(12), 3336–3348.

Benhamou, S. (2007). How many animals really do the levy walk? *Ecology*, *88*(8), 1962–1969.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289–300.

Black, J. J., & Mejabi, O. (1995). Simulation of complex manufacturing equipment reliability using object oriented methods. *Reliability Engineering & System Safety*, *48*(1), 11–18.

Buchin, M., Kruckenberg, H., & Kölzsch, A. (2013). Segmenting trajectories by movement states. In *Advances in spatial data handling* (pp. 15–25). Springer.

Bullard, F. (1999). *Estimating the home range of an animal: a brownian bridge approach. johns hopkins university* (Unpublished doctoral dissertation). Master thesis.

Burt, W. H. (1943). Territoriality and home range concepts as applied to mammals. *Journal of mammalogy*, *24*(3), 346–352.

Byrne, R., Noser, R., Bates, L., & Jupp, P. (2009). How did they get here from there? detecting changes of direction in terrestrial ranging. *Animal Behaviour*, *77*(3), 619–631.

Calenge, C. (2006). The package adehabitat for the r software: a tool for the analysis of space and habitat use by animals. *Ecological modelling*, *197*(3-4), 516–519.

Chang, Y. S., & Bai, D. S. (2001). Control charts for positively-skewed populations with weighted standard deviations. *Quality and Reliability Engineering International*, *17*(5), 397–406.

Chua, B., Coenen, E., Van Delen, J., & Weary, D. (2002). Effects of pair versus individual housing on the behavior and performance of dairy calves. *Journal of dairy science*, *85*(2), 360–364.

Clark, J. S., & Bjørnstad, O. N. (2004). Population time series: process variability, observation errors, missing values, lags, and hidden states. *Ecology*, *85*(11), 3140–3150.

Codling, E. A., & Plank, M. J. (2011). Turn designation, sampling rate and the misidentification of power laws in movement path data using maximum likelihood estimates. *Theoretical Ecology*, *4*(3), 397–406.

Couzin, I. D., Krause, J., Franks, N. R., & Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature*, *433*(7025), 513–516.

Dexter, F., Ledolter, J., & Hindman, B. J. (2014). Bernoulli cumulative sum (cusum) control charts for monitoring of anesthesiologists performance in supervising anesthesia residents and nurse anesthetists. *Anesthesia & Analgesia*, *119*(3), 679–685.

Dicke, M., & Burrough, P. A. (1988). Using fractal dimensions for characterizing tortuosity of animal trails. *Physiological Entomology*, *13*(4), 393–398.

Diosdado, J. A. V., Barker, Z. E., Hodges, H. R., Amory, J. R., Croft, D. P., Bell, N. J., & Codling, E. A. (2015). Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. *Animal Biotelemetry*, *3*(1), 15.

Douglas, D. H., & Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *10*(2), 112–122.

Downs, J., Horner, M., Lamb, D., Loraamm, R. W., Anderson, J., & Wood, B. (2018). Testing time-geographic density estimation for home range analysis using an agent-based model of animal movement. *International Journal of Geographical Information Science*, 1–18.

Downs, J. A., Horner, M. W., & Tucker, A. D. (2011). Time-geographic density estimation for home range analysis. *Annals of GIS*, *17*(3), 163–171.

Fauchald, P., & Tveraa, T. (2003). Using first-passage time in the analysis of area-restricted search and habitat selection. *Ecology*, *84*(2), 282–288.

Ford, R. G., & Krumme, D. W. (1979). The analysis of space use patterns. *Journal of Theoretical Biology*, *76*(2), 125–155.

Franke, A., Caelli, T., Kuzyk, G., & Hudson, R. J. (2006). Prediction of wolf (canis lupus) kill-sites using hidden markov models. *Ecological modelling*, *197*(1), 237–246.

González, L., Tolkamp, B., Coffey, M., Ferret, A., & Kyriazakis, I. (2008). Changes in feeding behavior as possible indicators for the automatic monitoring of health disorders in dairy cows. *Journal of Dairy Science*, *91*(3), 1017–1028.

Hawkins, D. M. (1987). Self-starting cusum charts for location and scale. *The Statistician*, 299–316.

Hawkins, D. M., & Olwell, D. H. (1998). Cumulative Sum Control Charts and Charting for Quality Control. Springer, New York.

Hawkins, D. M., & Olwell, D. H. (2012). *Cumulative sum charts and charting for quality improvement.* Springer Science & Business Media.

Hills, T. T., Kalff, C., & Wiener, J. M. (2013). Adaptive lévy processes and area-restricted search in human foraging. *PLoS One*, *8*(4), e60488.

Horne, J. S., Garton, E. O., Krone, S. M., & Lewis, J. S. (2007). Analyzing animal movements using brownian bridges. *Ecology*, *88*(9), 2354–2363.

Howery, L. D., Provenza, F. D., Banner, R. E., & Scott, C. B. (1996). Differences in home range and habitat use among individuals in a cattle herd. *Applied Animal Behaviour Science*, *49*(3), 305–320.

Humphries, N. E., Queiroz, N., Dyer, J. R., Pade, N. G., Musyl, M. K., Schaefer, K. M., … others (2010). Environmental context explains lévy and brownian movement patterns of marine predators. *Nature*, *465*(7301), 1066–1069.

Humphries, N. E., Weimerskirch, H., Queiroz, N., Southall, E. J., & Sims, D. W. (2012). Foraging success of biological lévy flights recorded in situ. *Proceedings of the National Academy of Sciences*, *109*(19), 7169–7174.

Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., ... Tsai, T. T. (2005). An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters*, *12*(2), 105–108.

James, A., Plank, M. J., & Edwards, A. M. (2011). Assessing lévy walks as models of animal foraging. *Journal of the Royal Society Interface*, *8*(62), 1233–1247.

Jennrich, R., & Turner, F. (1969). Measurement of non-circular home range. *Journal of Theoretical Biology*, *22*(2), 227–237.

Jonsen, I. D., Myers, R. A., & James, M. C. (2007). Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. *Marine Ecology Progress Series*, *337*, 255–264.

Jukan, A., Masip-Bruin, X., & Amla, N. (2017). Smart computing and sensing technologies for animal welfare: A systematic review. *ACM Computing Surveys (CSUR)*, *50*(1), 10.

Kapota, D., Dolev, A., & Saltz, D. (2017). Inferring detailed space use from movement paths: A unifying, residence time-based framework. *Ecology and Evolution*, *7*(20), 8507–8514.

Kays, R., Crofoot, M. C., Jetz, W., & Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science*, *348*(6240), aaa2478.

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

Knell, A. S., & Codling, E. A. (2012). Classifying area-restricted search (ars) using a partial sum approach. *Theoretical Ecology*, *5*(3), 325–339.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., & Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: hidden markov models and extensions. *Ecology*, *93*(11), 2336–2342.

Larson-Praplan, S., George, M., Buckhouse, J., & Laca, E. (2015). Spatial and temporal domains of scale of grazing cattle. *Animal Production Science*, *55*(3), 284–297.

López-López, P. (2016). Individual-based tracking systems in ornithology: welcome to the era of big data. *Ardeola*.

Macdonald, D., Ball, F., Hough, N., & Amlaner, C. (1980). The evaluation of home range size and configuration using radio tracking data. *A handbook on bio-telemetry and radio tracking*, 405–424.

Miller, K., & Wood-Gush, D. (1991). Some effects of housing on the social behaviour of dairy cows. *Animal Science*, *53*(3), 271–278.

Montgomery, D. C., & Gerth, R. (1998). Introduction to statistical quality control. *IIE Transactions*, *30*(6), 571.

Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., & Fryxell, J. M. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, *85*(9), 2436–2445.

Moreau, M., Siebert, S., Buerkert, A., & Schlecht, E. (2009). Use of a tri-axial accelerometer for automated recording and classification of goats grazing behaviour. *Applied Animal Behaviour Science*, *119*(3), 158–170.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, *105*(49), 19052–19059.

Norring, M., Häggman, J., Simojoki, H., Tamminen, P., Winckler, C., & Pastell, M. (2014). Lameness impairs feeding behavior of dairy cows. *Journal of dairy science*, *97*(7), 4317–4321.

Omnisense. (2013). Omnisense Series 500 Cluster Geolocation System Data Sheet. `http://omnisense.co.uk/assets/omspd004_s500systemdatasheet_0-4.pdf`. (Accessed: 06-09-2017)

Page, E. (1954). Continuous inspection schemes. *Biometrika*, *41*(1/2), 100–115.

Pastell, M., Hietaoja, J., Yun, J., Tiusanen, J., & Valros, A. (2016). Predicting farrowing of sows housed in crates and pens using accelerometers and cusum charts. *Computers and Electronics in Agriculture*, *127*, 197–203.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., & Matthiopoulos, J. (2008). State–space models of individual animal movement. *Trends in ecology & evolution*, *23*(2), 87–94.

Pazhayamadom, D. G., Kelly, C. J., Rogan, E., & Codling, E. A. (2013). Self-starting cusum approach for monitoring data poor fisheries. *Fisheries Research*, *145*, 114–127.

Pazhayamadom, D. G., Kelly, C. J., Rogan, E., & Codling, E. A. (2015). Decision interval cumulative sum harvest control rules (di-cusum-hcr) for managing fisheries with limited historical information. *Fisheries Research*, *171*, 154–169.

Pencina, M. J., D'Agostino, R. B., & Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: from area under the roc curve to reclassification and beyond. *Statistics in medicine*, *27*(2), 157–172.

Potter, M., & Broom, D. (1987). The behaviour and welfare of cows in relation to cubicle house design. *Cattle housing systems, lameness and behaviour*, 129–147.

Proudfoot, K., Jensen, M. B., Heegaard, P. M., & von Keyserlingk, M. (2013). Effect of moving dairy cows at different stages of labor on behavior during parturition. *Journal of dairy science*, *96*(3), 1638–1646.

Pyke, G. H. (2015). Understanding movements of organisms: it's time to abandon the lévy foraging hypothesis. *Methods in Ecology and Evolution*, *6*(1), 1–16.

Reiczigel, J., Zakariás, I., & Rózsa, L. (2005). A bootstrap test of stochastic equality of two populations. *The American Statistician*, *59*(2), 156–161.

Reynolds, A. (2014). Lévy flight movement patterns in marine predators may derive from turbulence cues. In *Proc. r. soc. a* (Vol. 470, p. 20140408).

Robert, B., White, B., Renter, D., & Larson, R. (2009). Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. *Computers and Electronics in Agriculture*, *67*(1), 80–84.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC bioinformatics*, *12*(1), 77.

Robinson, P., Tremblay, Y., Crocker, D., Kappes, M., Kuhn, C., Shaffer, S., ... Costa, D. (2007). A comparison of indirect measures of feeding behaviour based on argos tracking data. *Deep Sea Research Part II: Topical Studies in Oceanography*, *54*(3), 356–368.

Seaman, D. E., & Powell, R. A. (1996). An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology*, *77*(7), 2075–2085.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611.

Sims, D. W., Humphries, N. E., Bradford, R. W., & Bruce, B. D. (2012). Lévy flight and brownian search patterns of a free-ranging predator reflect different prey field characteristics. *Journal of Animal Ecology*, *81*(2), 432–442.

Sims, D. W., Southall, E. J., Humphries, N. E., Hays, G. C., Bradshaw, C. J., Pitchford, J. W., ... others (2008). Scaling laws of marine predator search behaviour. *Nature*, *451*(7182), 1098–1102.

Stonehouse, J. M., & Forrester, G. J. (1998). Robustness of the t and u tests under combined assumption violations. *Journal of Applied Statistics*, *25*(1), 63–74.

Sueur, C. (2011). A non-lévy random walk in chacma baboons: what does it mean? *PloS one*, *6*(1), e16131.

Thiebault, A., & Tremblay, Y. (2013). Splitting animal trajectories into fine-scale behaviorally consistent movement units: breaking points relate to external stimuli in a foraging seabird. *Behavioral Ecology and Sociobiology*, *67*(6), 1013–1026.

Tomkiewicz, S. M., Fuller, M. R., Kie, J. G., & Bates, K. K. (2010). Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *365*(1550), 2163–2176.

Travis, J. (2008). How a shark finds its next meal. *ScienceNOW Daily News*.

Tuckwell, H. C., & Wan, F. Y. (1984). First-passage time of markov processes to moving barriers. *Journal of applied probability*, *21*(4), 695–709.

Van Winkle, W. (1975). Comparison of several probabilistic home-range models. *The Journal of wildlife management*, 118–123.

Visvalingam, M., & Whyatt, J. (1992). Line generalisation by repeated elimination of the smallest area. *The Cartographic Journal*, *30*(1), 46–51.

Viswanathan, G. M., Afanasyev, V., Buldyrev, S. V., Havlin, S., Da Luz, M., Raposo, E., & Stanley, H. E. (2000). Lévy flights in random searches. *Physica A: Statistical Mechanics and its Applications*, *282*(1), 1–12.

Viswanathan, G. M., Buldyrev, S. V., Havlin, S., Da Luz, M., Raposo, E., & Stanley, H. E. (1999). Optimizing the success of random searches. *Nature*, *401*(6756), 911–914.

Voegeli, F. A., Smale, M. J., Webber, D. M., Andrade, Y., & ODor, R. K. (2001). Ultrasonic telemetry, tracking and automated monitoring technology for sharks. In *The behavior and sensory biology of elasmobranch fishes: an anthology in memory of donald richard nelson* (pp. 267–282). Springer.

Weimerskirch, H., Pinaud, D., Pawlowski, F., & Bost, C.-A. (2007). Does prey capture induce area-restricted search? a fine-scale study using gps in a marine predator, the wandering albatross. *The American Naturalist*, *170*(5), 734–743.

Welch, B. L. (1947). The generalization ofstudent's' problem when several different population variances are involved. *Biometrika*, *34*(1/2), 28–35.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, *1*(6), 80–83.

Woillez, M., Fablet, R., Ngo, T.-T., Lalire, M., Lazure, P., & De Pontual, H. (2016). A hmm-based model to geolocate pelagic fish from high-resolution individual temperature and depth histories: European sea bass as a case study. *Ecological Modelling*, *321*, 10–22.

Woor, A. (2017). Forming and applying a novel hybrid CUSUM to determine differences in lame and non-lame dairy cows.

Worton, B. J. (1989). Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*, *70*(1), 164–168.

Zaburdaev, V., Denisov, S., & Klafter, J. (2015). Lévy walks. *Reviews of Modern Physics*, *87*(2), 483.

Zhao, K., & Jurdak, R. (2016). Understanding the spatiotemporal pattern of grazing cattle movement. *Scientific reports*, *6*, 31967.

Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, *57*(1), 173–181.

A Initial method for detecting when the system is back in-control

The initial method to detect when the system is back in-control involved analysing the DFO data and recording when successive distances had become acceptably close together. If the Hybrid CUSUM detected the system to go out-of-control at point j, the mean of the x and y ordinates whilst the system was in control was recorded as the most-recent residency position (x_0, y_0) and all position coordinates after this were converted into one-dimensional data by measuring the distance from this most-recent residency position:

$$\text{Distance From Origin at point } i = D_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}$$

where (x_0, y_0) *is the location of the origin and* (x_i, y_i) *is the location of the animal at a time step* i

An arbitrary number n of these successive DFOs were considered and their absolute differences normalised and checked to see if they were within a given tolerance level t:

$$\begin{aligned}
\frac{D_{i+1} - D_i}{\sigma} &< t \\
\frac{D_{i+2} - D_{i+1}}{\sigma} &< t \\
&\vdots \\
\frac{D_n - D_{n-1}}{\sigma} &< t
\end{aligned}$$

If enough of these successive differences were within the tolerance level t then the system was thought to be in-control and the Hybrid CUSUM could restart.

This initial method to detect when the system is back in-control was used and the detected residency positions were plotted over the position trajectory (as in Figure 9) to verify if the algorithm was correctly identifying residencies. Whilst most obvious residencies were detected, some trajectories which clearly should correspond to transitions were also incorrectly detected as residencies.

Efforts were made to change the number of successive distances n that were considered and the number of these successive distances that were required to be within the tolerance level t. Whilst this eliminated some false-positive residencies, it did not eliminate all. The point at which the system was detected as being back in-control was also varied; for the window size of n distances that were being considered, would the system be defined as back in-control at the start of the window, the end of the window or at some point in-between? By varying this back in-control point it was found that the end of the window seemed to give the least false-positive residency detections, but again, did not eliminate all.

Upon further investigation, it was realised that the false-positive residency detection was mainly where the cow turned round and changed direction. This highlighted that this method had the same disadvantage as that shown in Figure 4a; since the distance of each point whilst out-of-control was being measured from the most-recent residency position (x_0, y_0) which meant that when the cow was changing direction (but still in a transition), the distances from (x_0, y_0) would not change by a great deal. As such, this method would falsely-detect that the cow was remaining in the same position, when in fact it was simply changing the direction of its trajectory.

The improved method that was used was to measure the distances of each of the n points from the average position of these n points (see Section 3.5.2). This effectively provided a shifting origin (x_0, y_0) from which to calculate the distance from and would focus more on the clustering of the points.

B Results of the Welch t-Test

Although the Wilcoxon Signed Rank Test was used in the main analysis, the results for the same data using the Welch t-Test is displayed here to compare the p-values of each test and to provide an increased level of confidence in the results of the Wilcoxon Signed Rank Test. Notice that the same categories were flagged as significant using the Welch t-Test as were flagged as significant using the Wilcoxon Signed Rank Test.

Out-of-control window size of 11 time points					
	t-value	**p-value**	**Lame Mean Value**	**Non-Lame Mean Value**	**Significant Difference in favour of**
Number of residencies	-0.12	0.9021	39.3	39.5	
Number of transitions	-0.12	0.9029	38.3	38.5	
Duration of resting residencies	3.66	0.0004	5985.4	5315.9	Lame
Duration of feeding residencies	-3.14	0.0025	2153.8	2553.1	Non-Lame
Duration of milking residencies	1.07	0.2880	1121.2	1011.6	
Number of resting residencies	-1.22	0.2252	14.0	15.7	
Number of feeding residencies	-3.76	0.0003	14.9	17.8	Non-Lame
Number of milking residencies	0.40	0.6933	10.4	10.0	
Resting to resting transitions	-1.12	0.2662	7.1	8.5	
Resting to feeding transitions	-3.15	0.0022	3.8	4.6	Non-Lame
Resting to milking transitions	2.25	0.0271	2.2	1.8	Lame
Feeding to resting transitions	-0.44	0.6642	5.4	5.5	
Feeding to feeding transitions	-3.33	0.0013	8.6	11.0	Non-Lame
Feeding to milking transitions	-1.45	0.1501	0.8	1.0	
Milking to resting transitions	-0.33	0.7421	0.9	1.0	
Milking to feeding transitions	1.31	0.1937	2.3	2.1	
Milking to milking transitions	0.22	0.8271	7.1	6.9	

Table 5: Results of the Welch t-test for an out-of-control window size of 11 time points carried out at the 5% level of significance.

Out-of-control window size of 75 time points					
	t-value	**p-value**	**Lame Mean Value**	**Non-Lame Mean Value**	**Significant Difference in favour of**
Number of residencies	-0.03	0.9749	21.7	21.7	
Number of transitions	-0.05	0.9576	20.8	20.9	
Duration of resting residencies	3.76	0.0003	5646.3	4901.1	Lame
Duration of feeding residencies	-2.71	0.0083	1651.8	2025.1	Non-Lame
Duration of milking residencies	1.47	0.1459	570.8	451.5	
Number of resting residencies	0.44	0.6577	9.8	9.5	
Number of feeding residencies	-3.26	0.0016	8.3	10.0	Non-Lame
Number of milking residencies	0.71	0.4799	3.5	3.3	
Resting to resting transitions	0.67	0.5028	3.9	3.6	
Resting to feeding transitions	-1.25	0.2145	3.3	3.7	
Resting to milking transitions	0.43	0.6666	1.7	1.6	
Feeding to resting transitions	0.11	0.9146	4.3	4.3	
Feeding to feeding transitions	-3.82	0.0003	3.2	4.9	Non-Lame
Feeding to milking transitions	-0.27	0.7901	0.6	0.6	
Milking to resting transitions	-0.41	0.6862	0.9	1.0	
Milking to feeding transitions	0.93	0.3534	1.6	1.4	
Milking to milking transitions	0.77	0.4404	1.0	0.8	

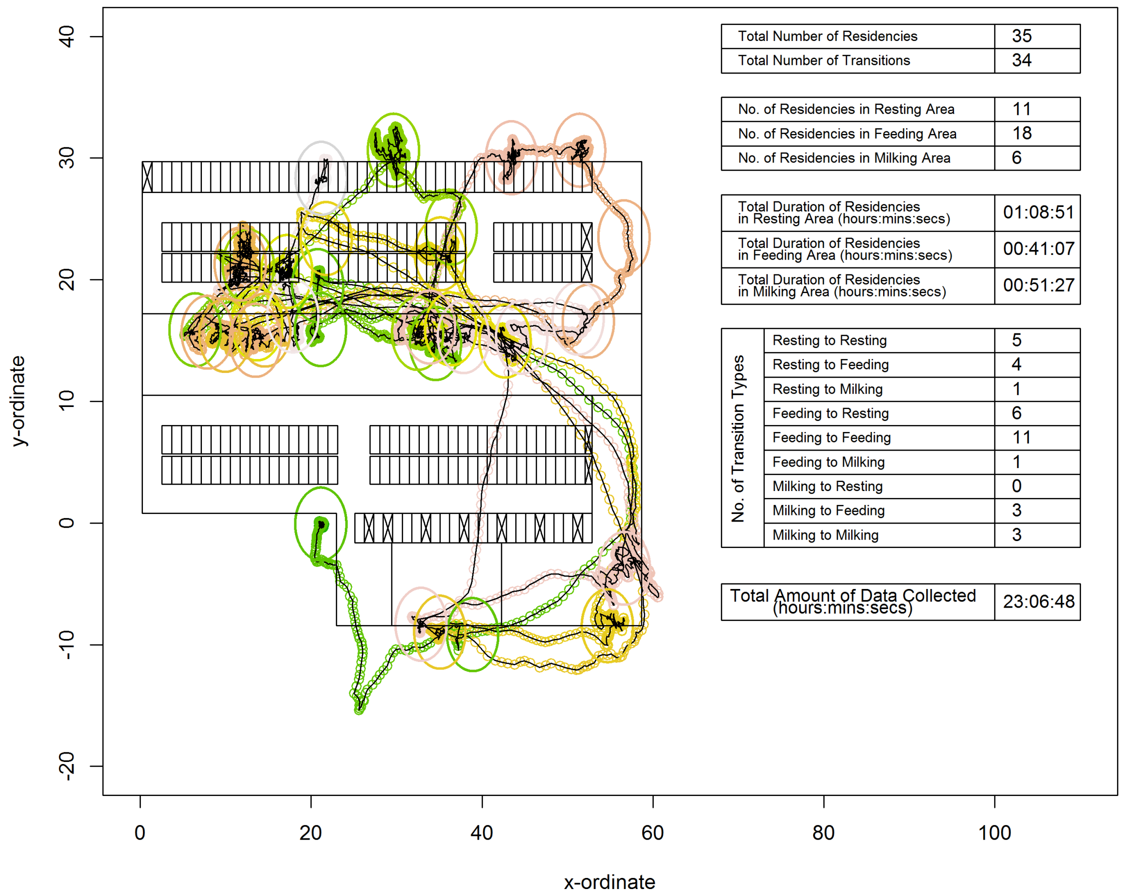
Table 6: Results of the Welch t-test for an out-of-control window size of 75 time points carried out at the 5% level of significance.

C Summary of cow data across days 2-6 with an out-of-control window size of 11 time points

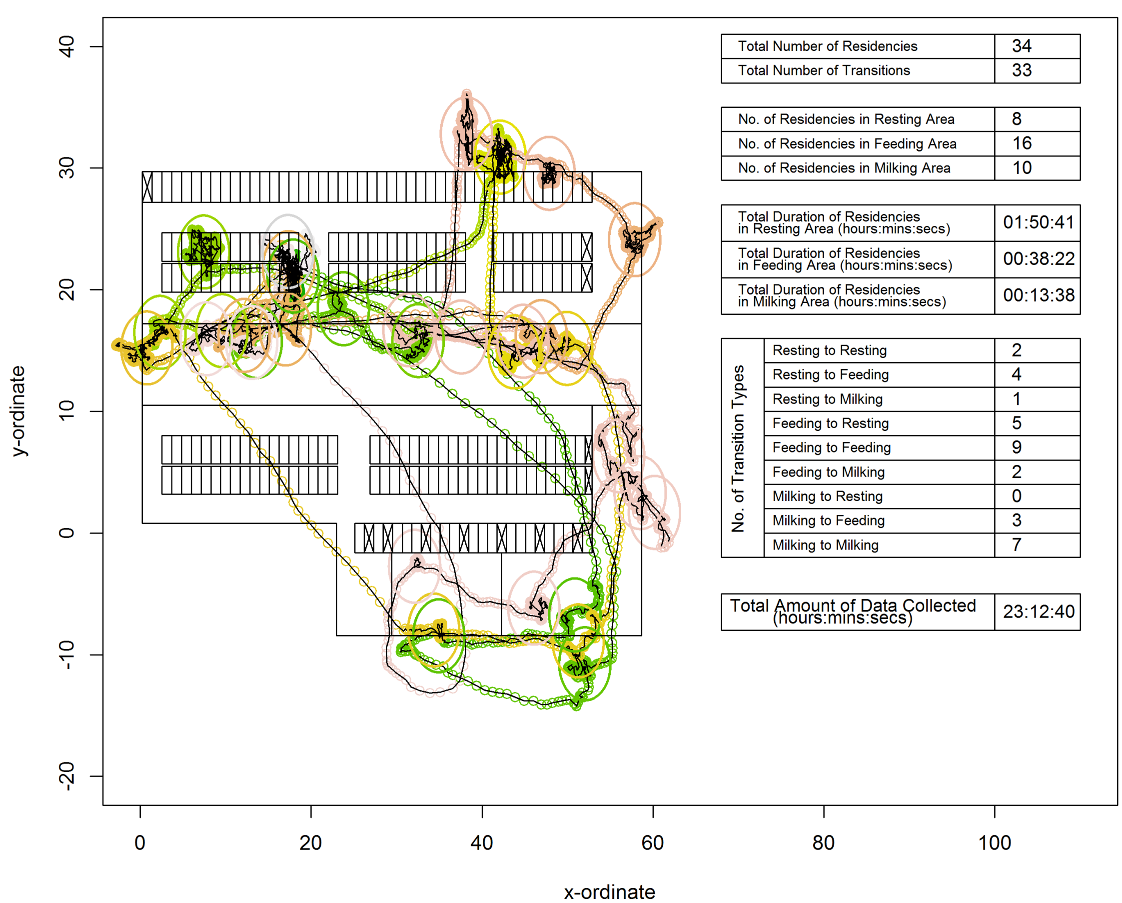
In order to summarise the data for each cow, the x and y position coordinates have been plotted and colour-coded to give a sense of time; green represents the start of the day, transitioning into yellow and then transitioning into pink to represent the end of the day. On the right side of the plot, a summary table has been plotted to show the results of each cow per day. All data given is using an out-of-control window size of 11 time points. The cow ID and day number are given at the top of each plot. Note that "Day 2" corresponds to the 23$^{\text{rd}}$ January and "Day 6" corresponds to the 27$^{\text{th}}$ January which gives the interval of days which were used for analysis in this report (see Section 4.1).

The position plots in this appendix have been superimposed over an accurate map of the barn which has been constructed using the relative coordinates. As such it is possible to visualise the different areas of the barn where the cow is situated and residencies are represented by the circles as discussed in Section 4.1.1. It can be observed that there are occasions when the cow is detected to be beyond the confines of the barn which is due to sensor error and noise in the system. Since the Hybrid CUSUM application aims to detect residencies, this was not a concern since if a residency is detected outside of the barn then it is probably the case that a cow is still in a residency state but the exact location of that residency is incorrect. As such, the decision was made to ignore these inaccuracies. Note that when analysis was carried out on the location of residencies, the inaccurate sensor locations were still within the required tolerance to be included in their "true" locations; for example when a cow was recorded to be above the resting area of the barn then it was probably the case that the cow was actually in the resting area and so values were chosen to included these inaccurate locations within the relevant resting, feeding or milking areas.

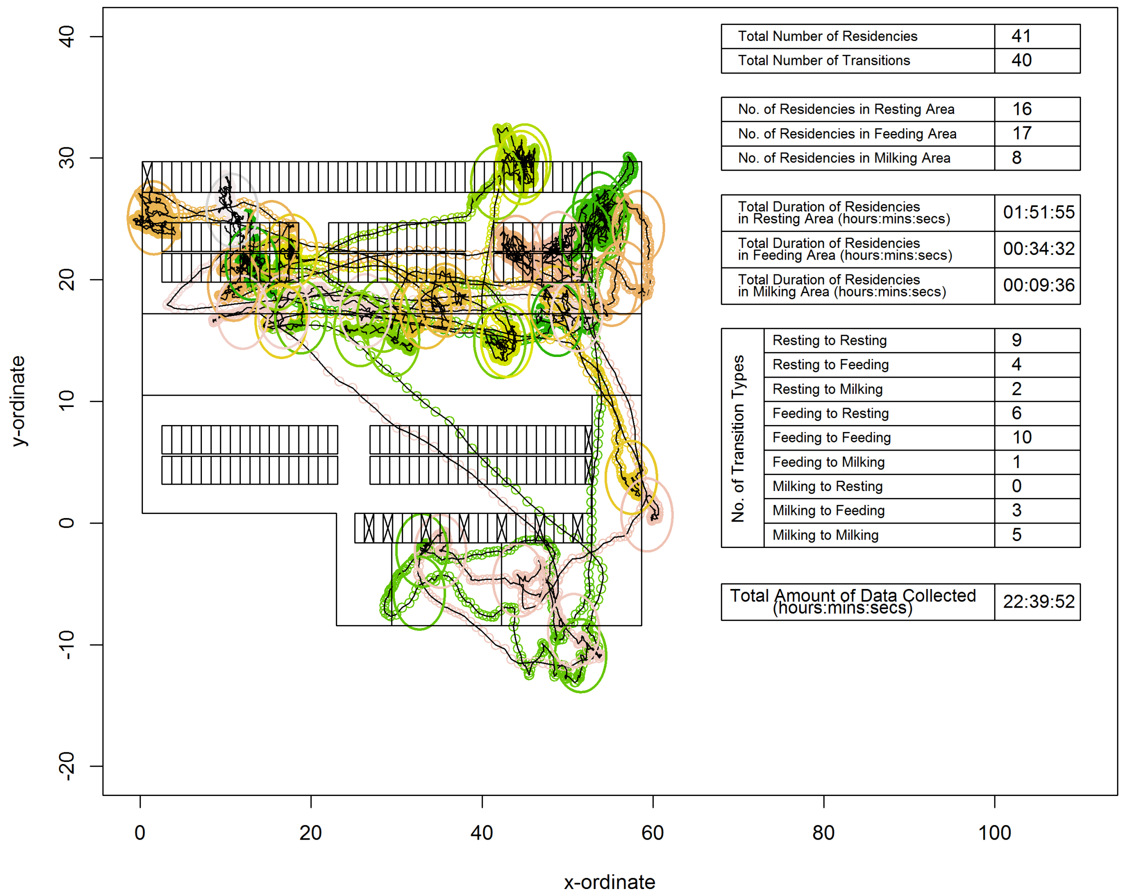
Cow 2344 Day 2 - Out-of-Control Window Size of 11 time points



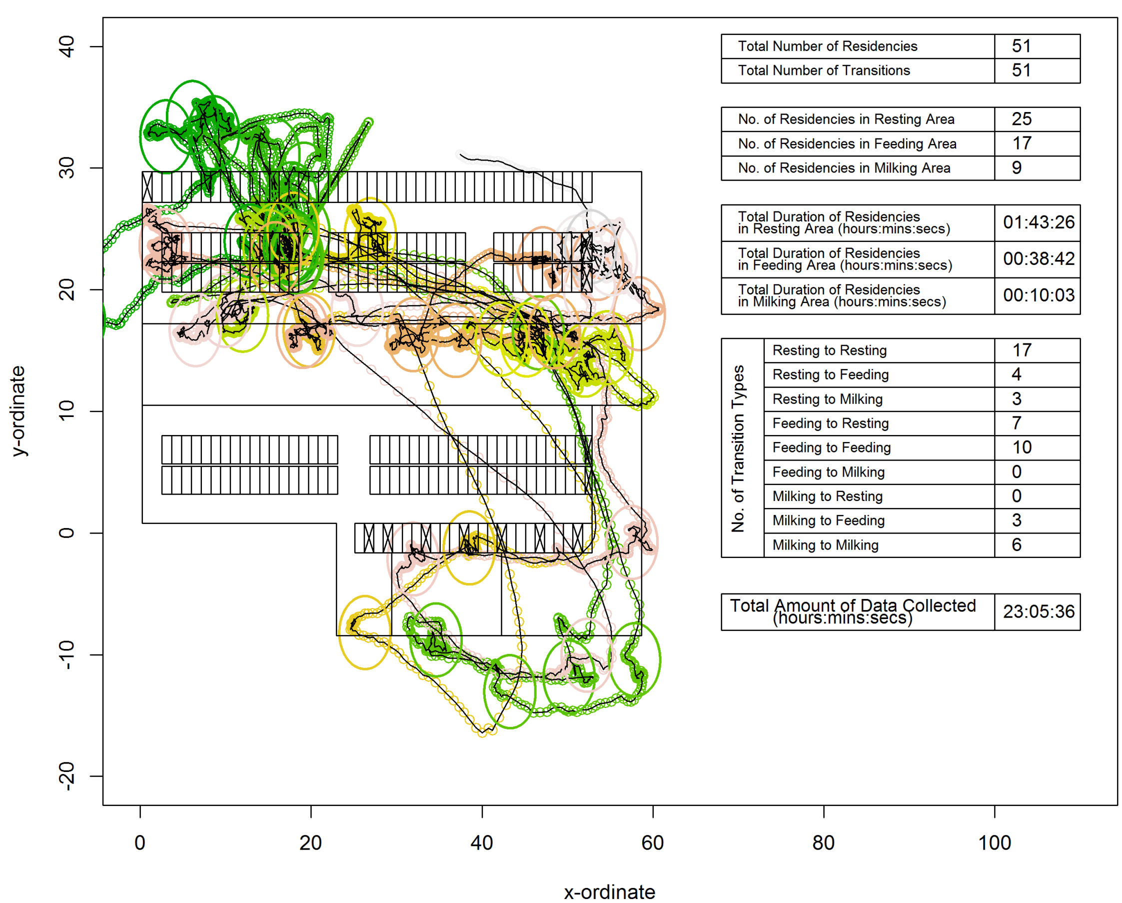
Cow 2344 Day 3 - Out-of-Control Window Size of 11 time points



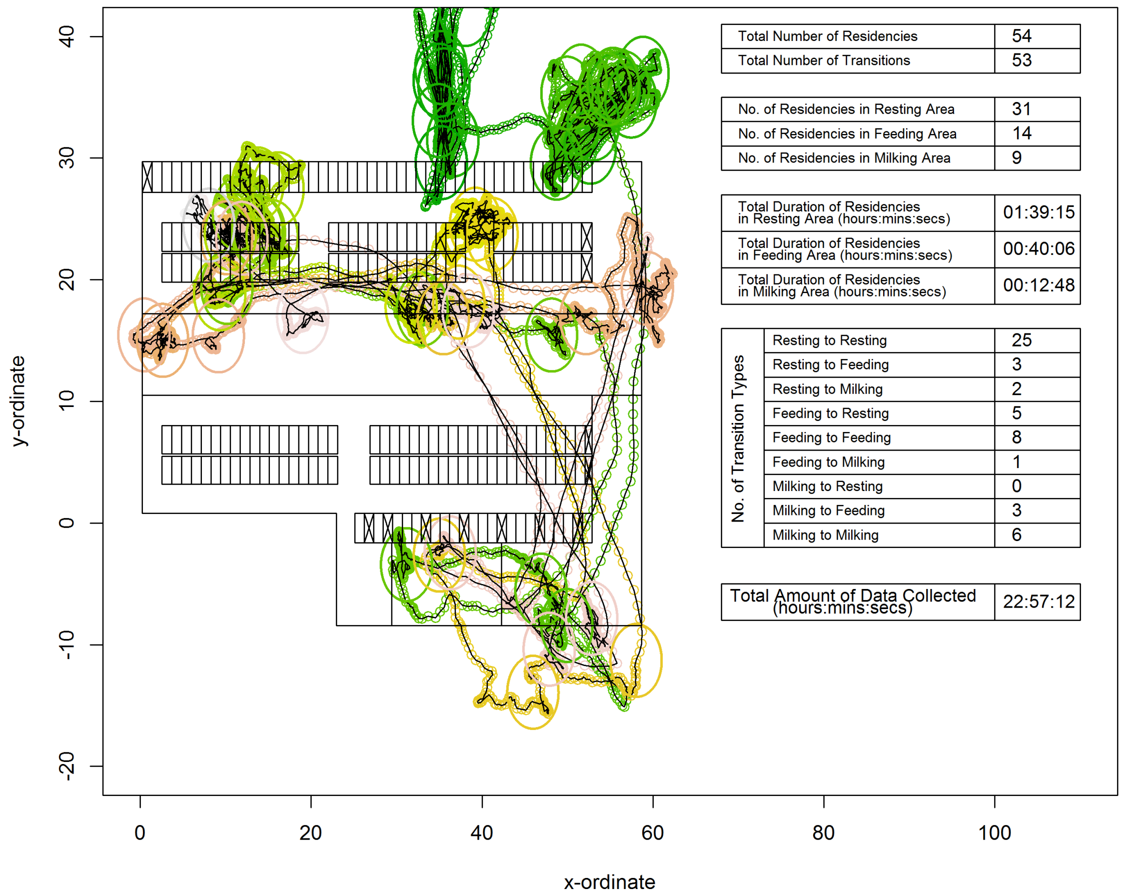
Cow 2344 Day 4 - Out-of-Control Window Size of 11 time points



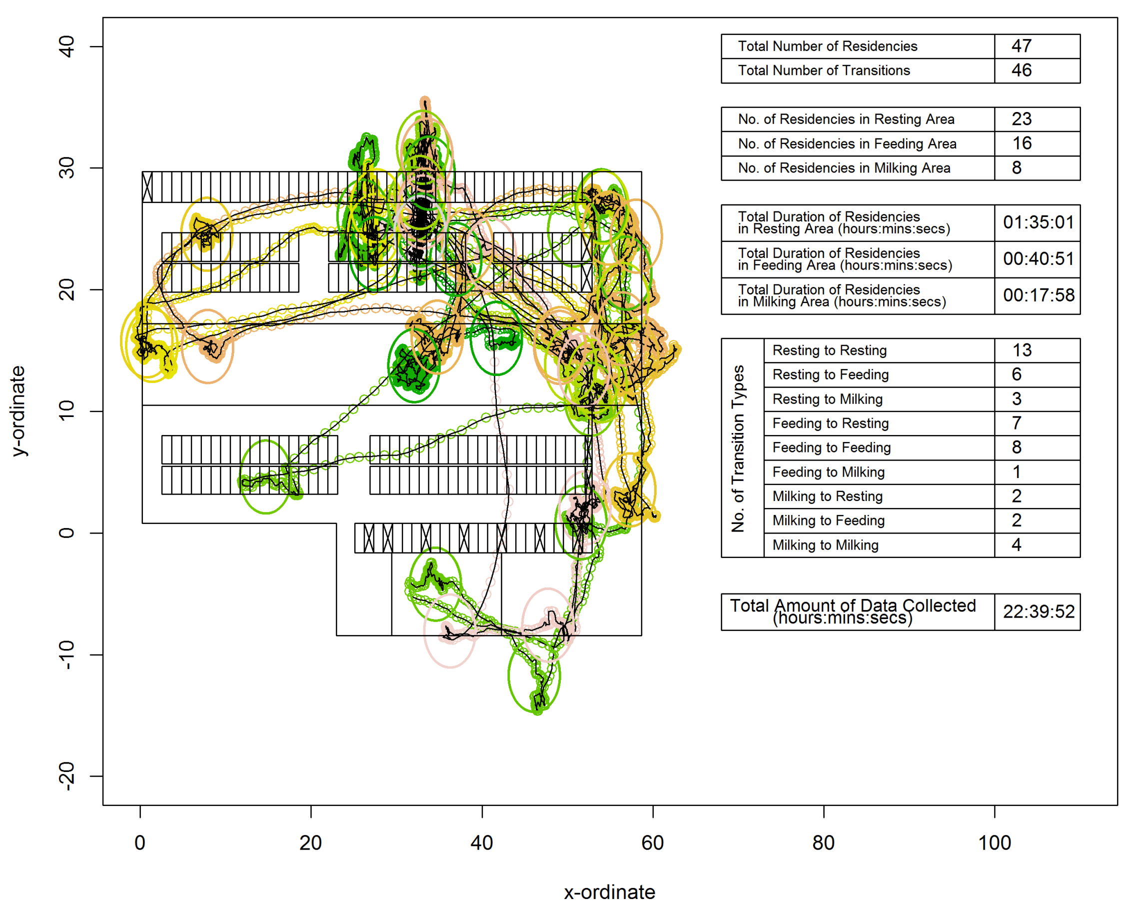
Cow 2344 Day 5 - Out-of-Control Window Size of 11 time points



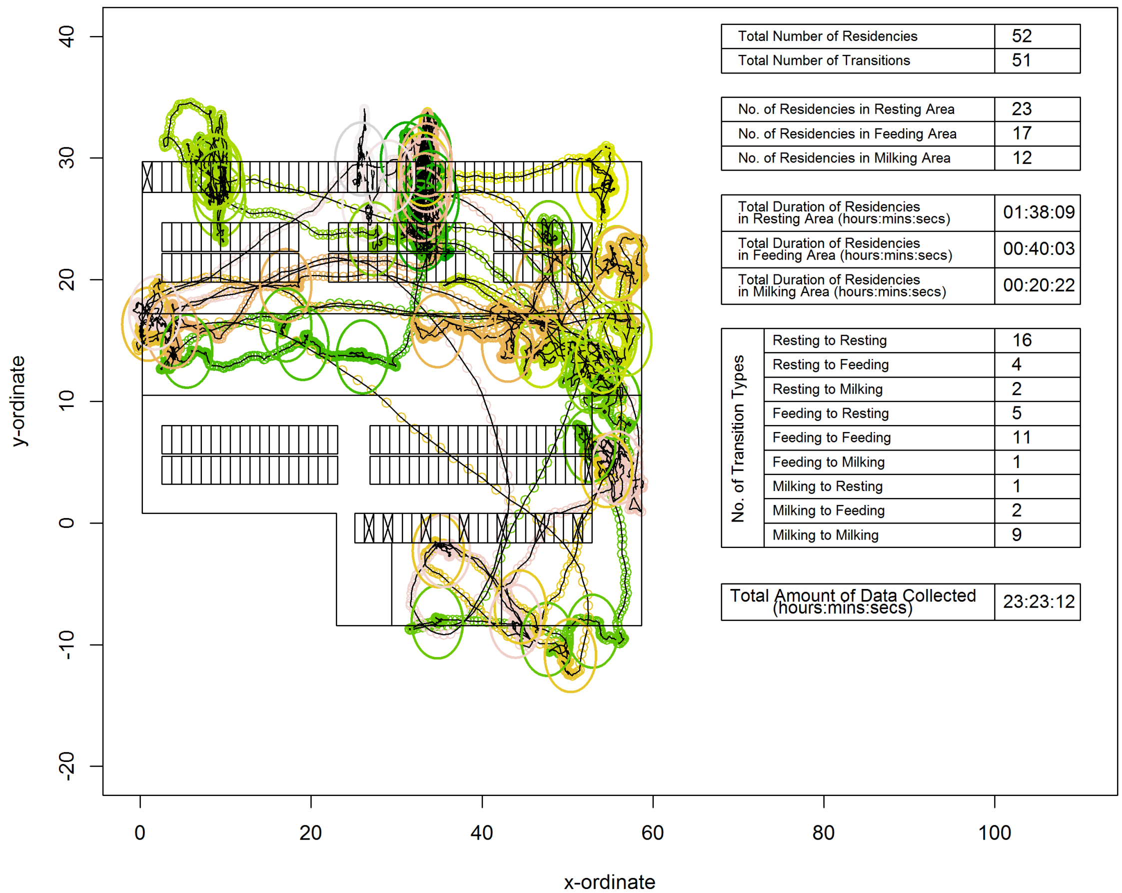
Cow 2344 Day 6 - Out-of-Control Window Size of 11 time points



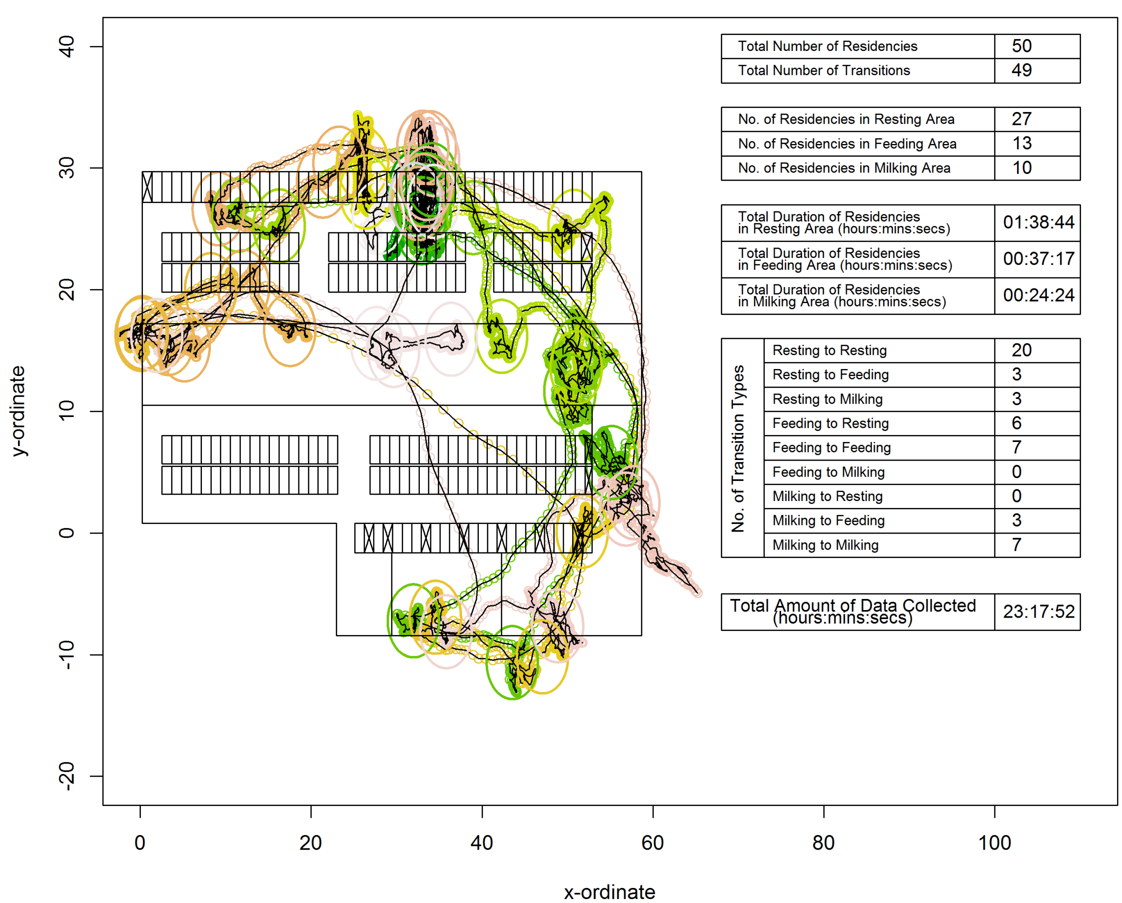
Cow 2010 Day 2 - Out-of-Control Window Size of 11 time points



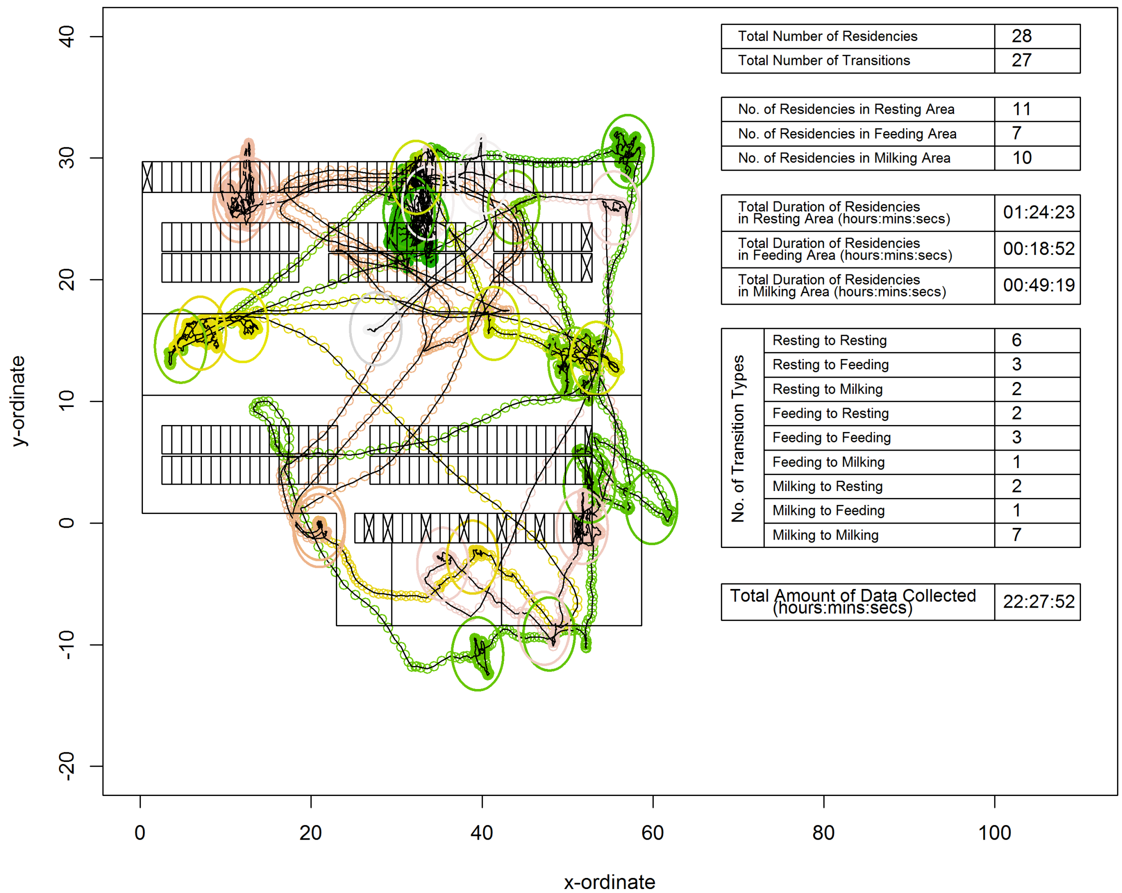
Cow 2010 Day 3 - Out-of-Control Window Size of 11 time points



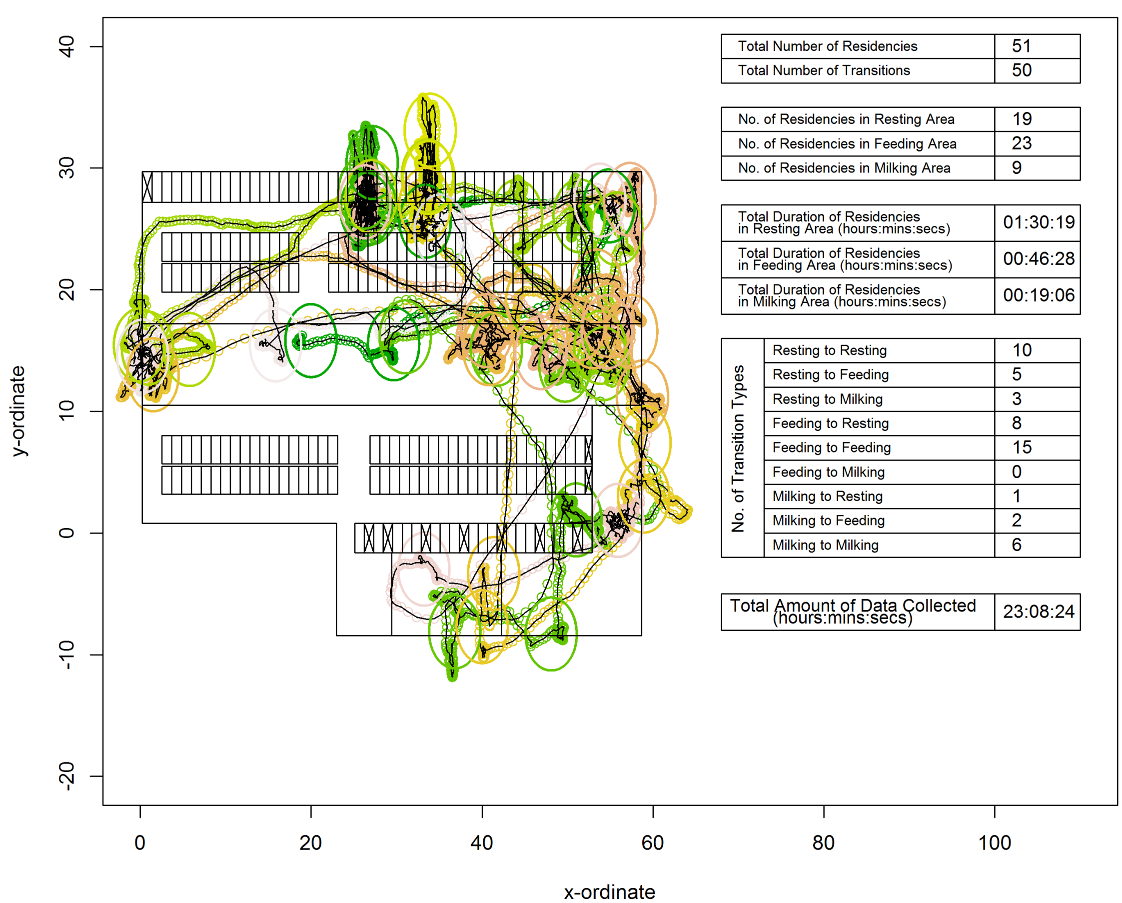
Cow 2010 Day 4 - Out-of-Control Window Size of 11 time points



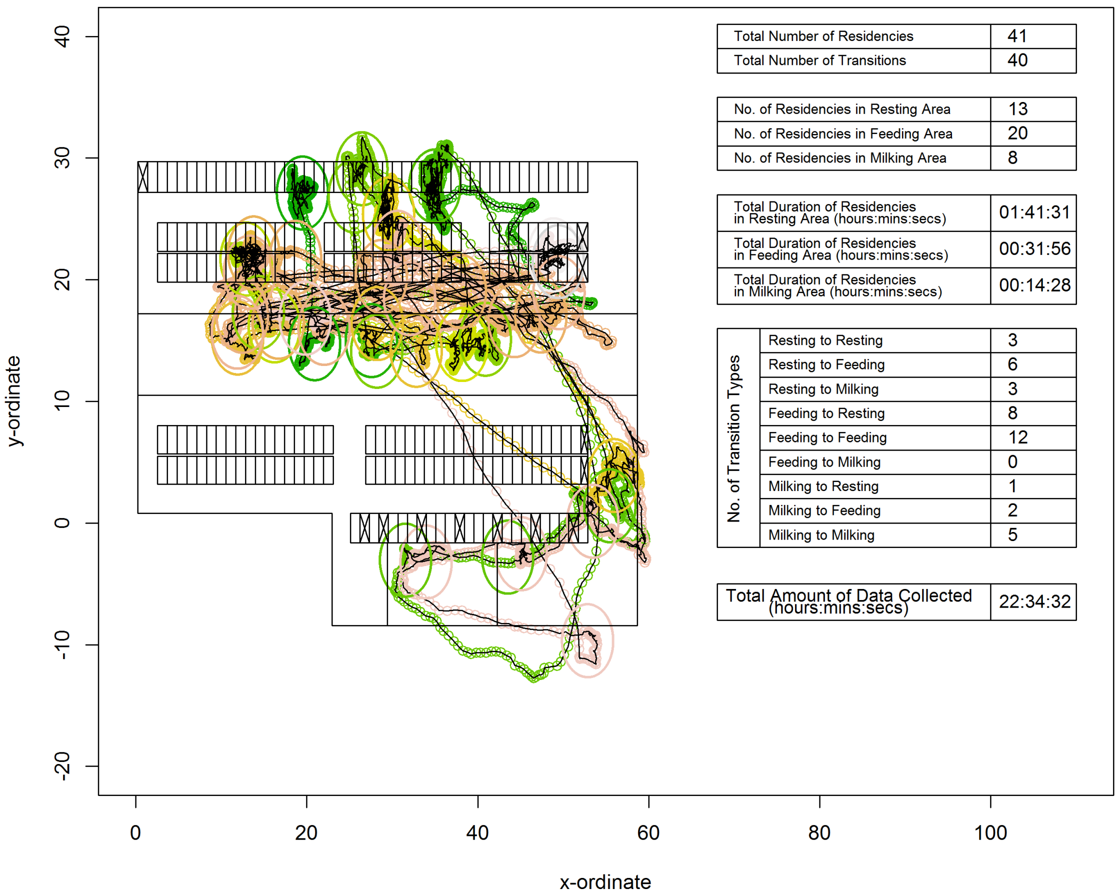
Cow 2010 Day 5 - Out-of-Control Window Size of 11 time points



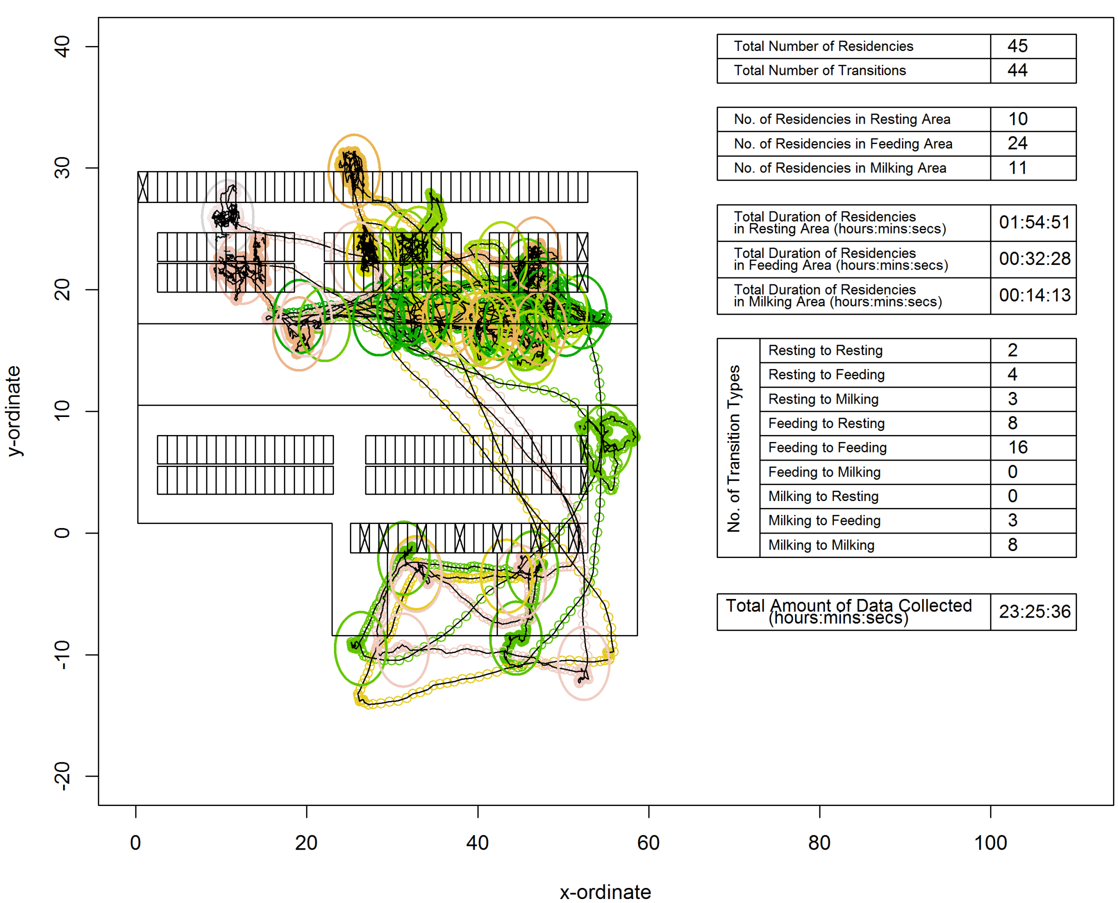
Cow 2010 Day 6 - Out-of-Control Window Size of 11 time points



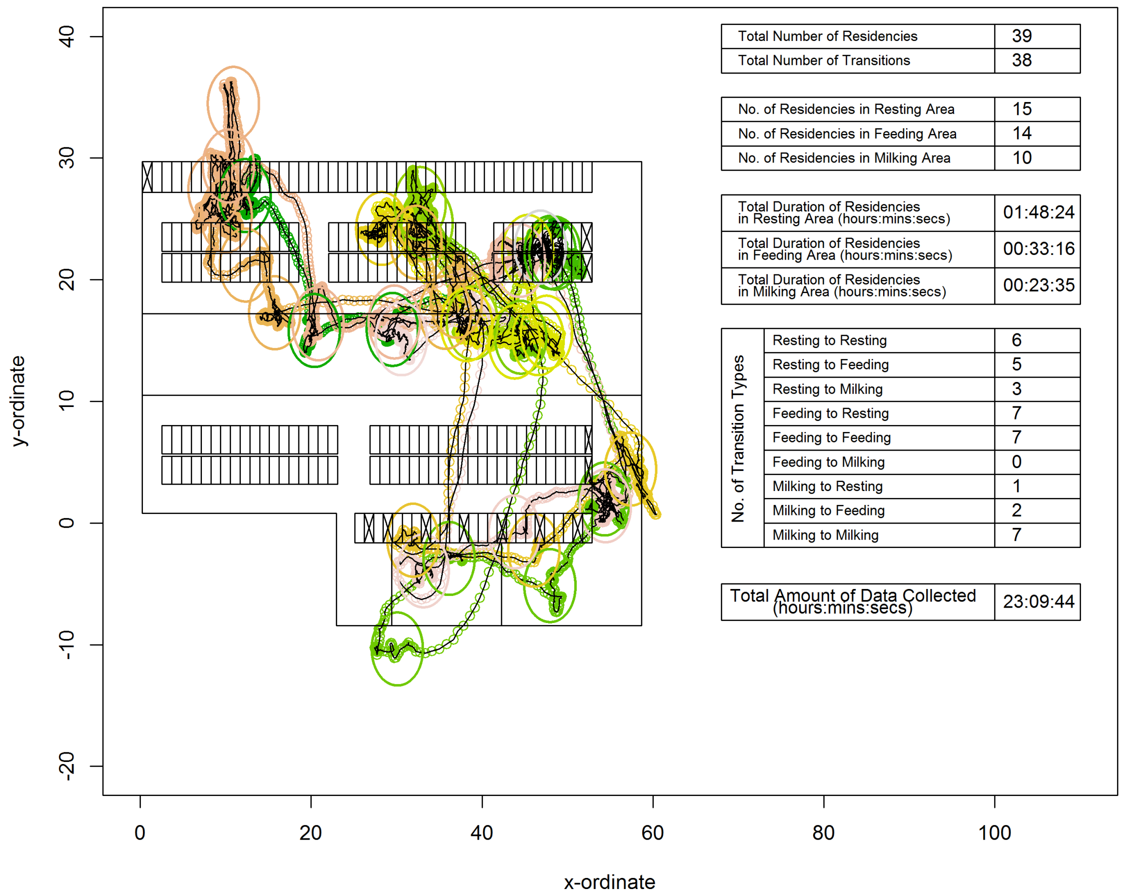
Cow 2060 Day 2 - Out-of-Control Window Size of 11 time points



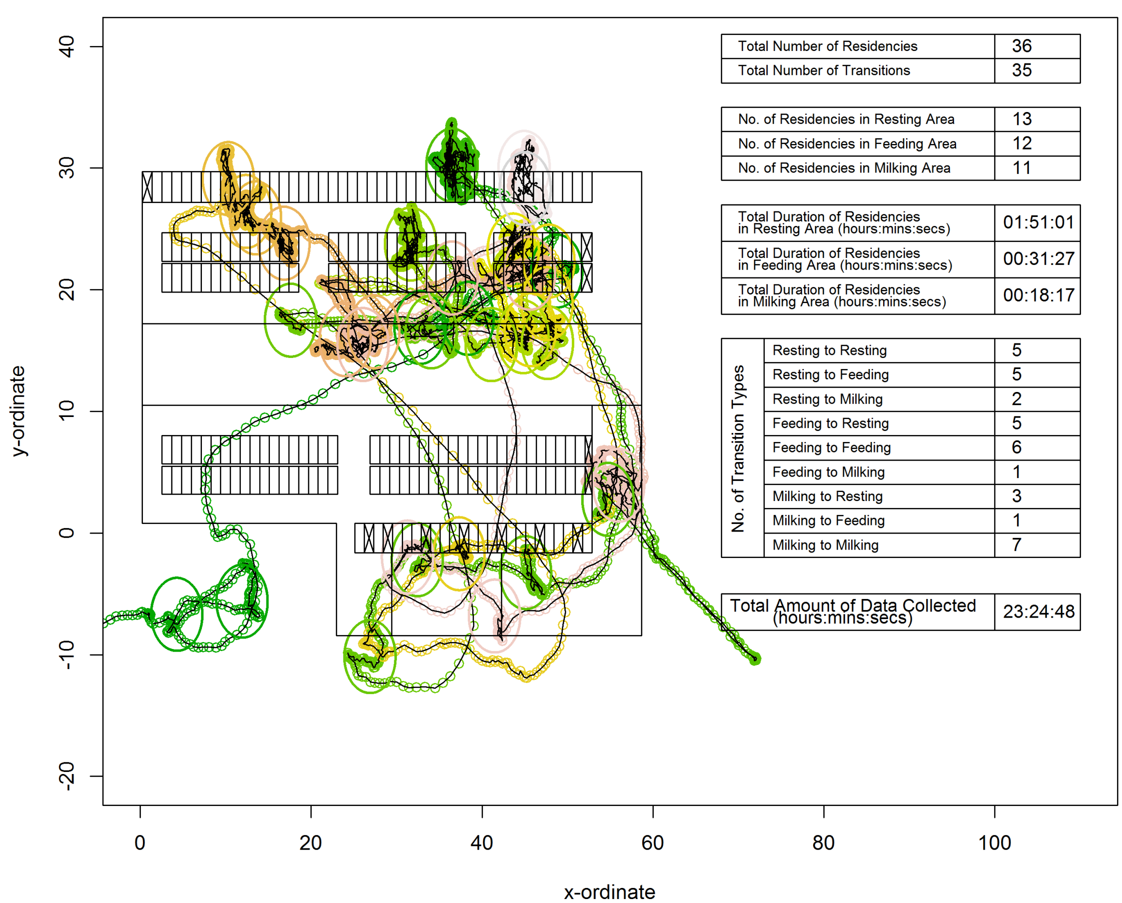
Cow 2060 Day 3 - Out-of-Control Window Size of 11 time points



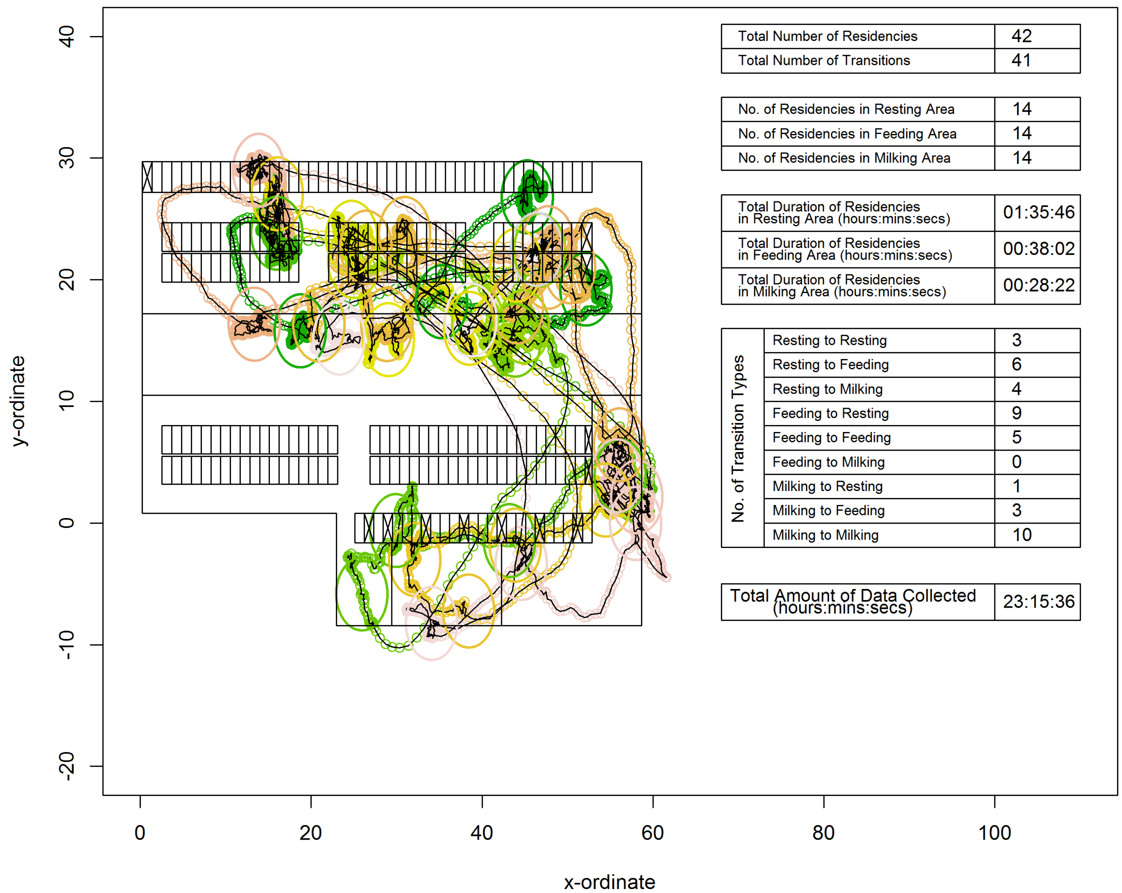
Cow 2060 Day 4 - Out-of-Control Window Size of 11 time points



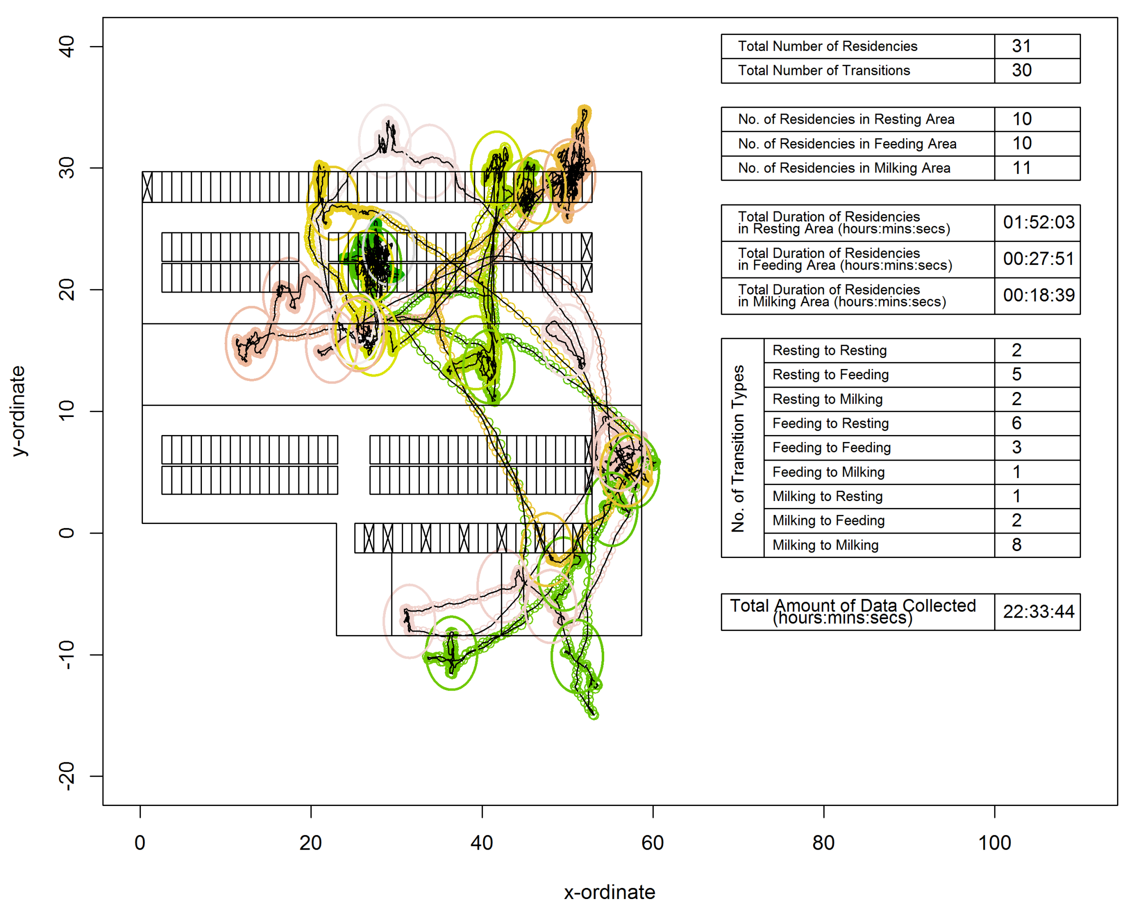
Cow 2060 Day 5 - Out-of-Control Window Size of 11 time points



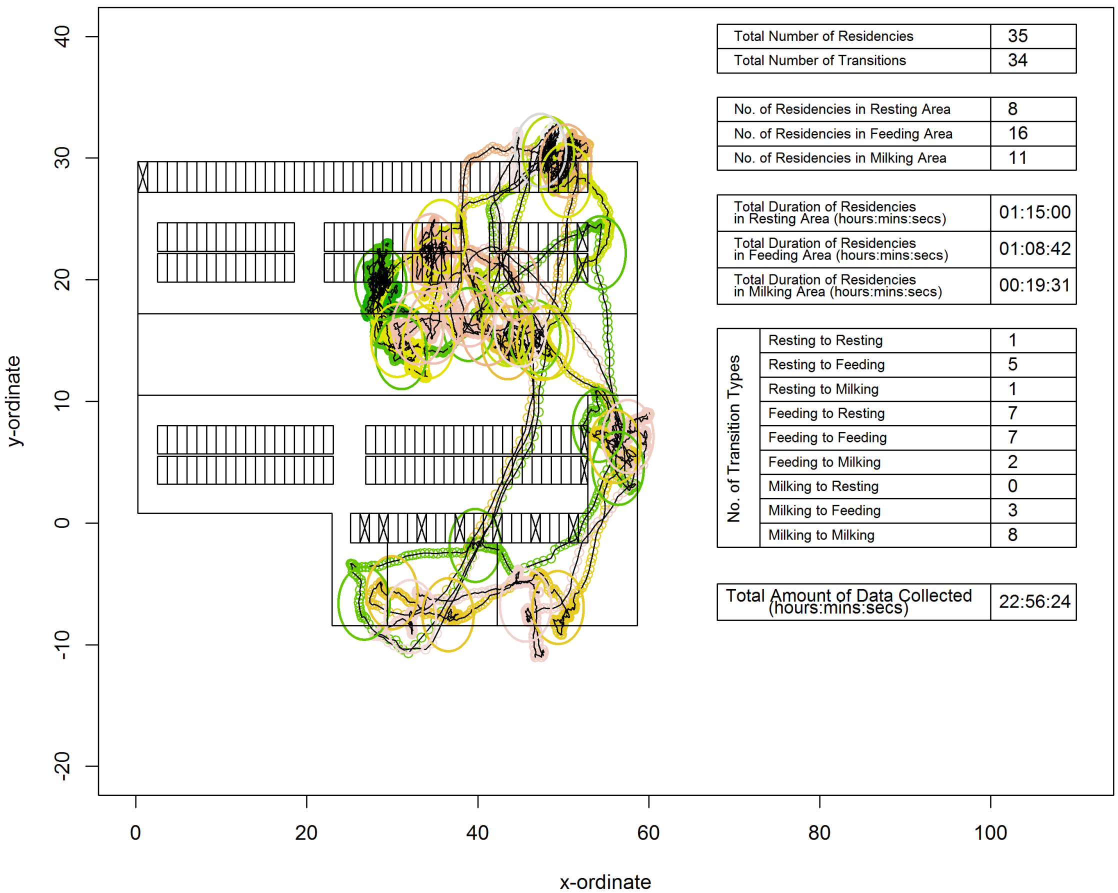
Cow 2060 Day 6 - Out-of-Control Window Size of 11 time points



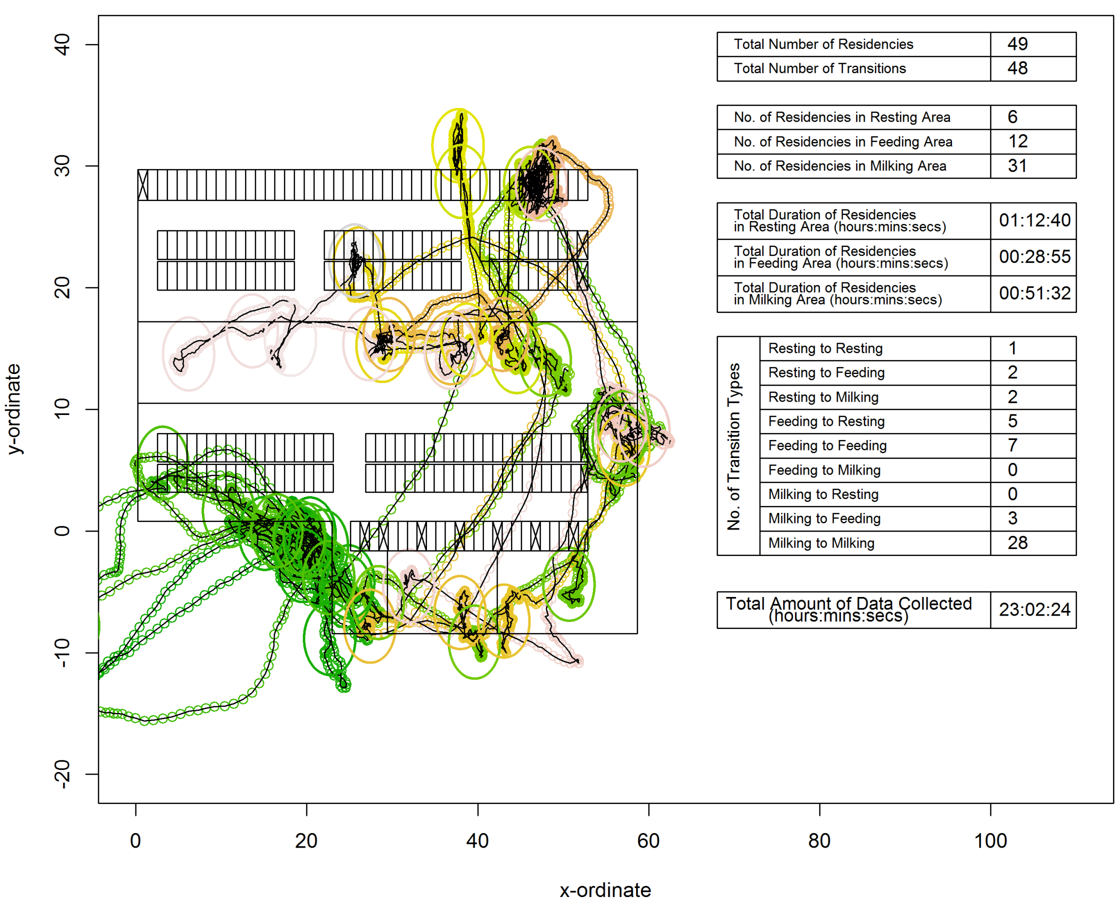
Cow 1891 Day 2 - Out-of-Control Window Size of 11 time points



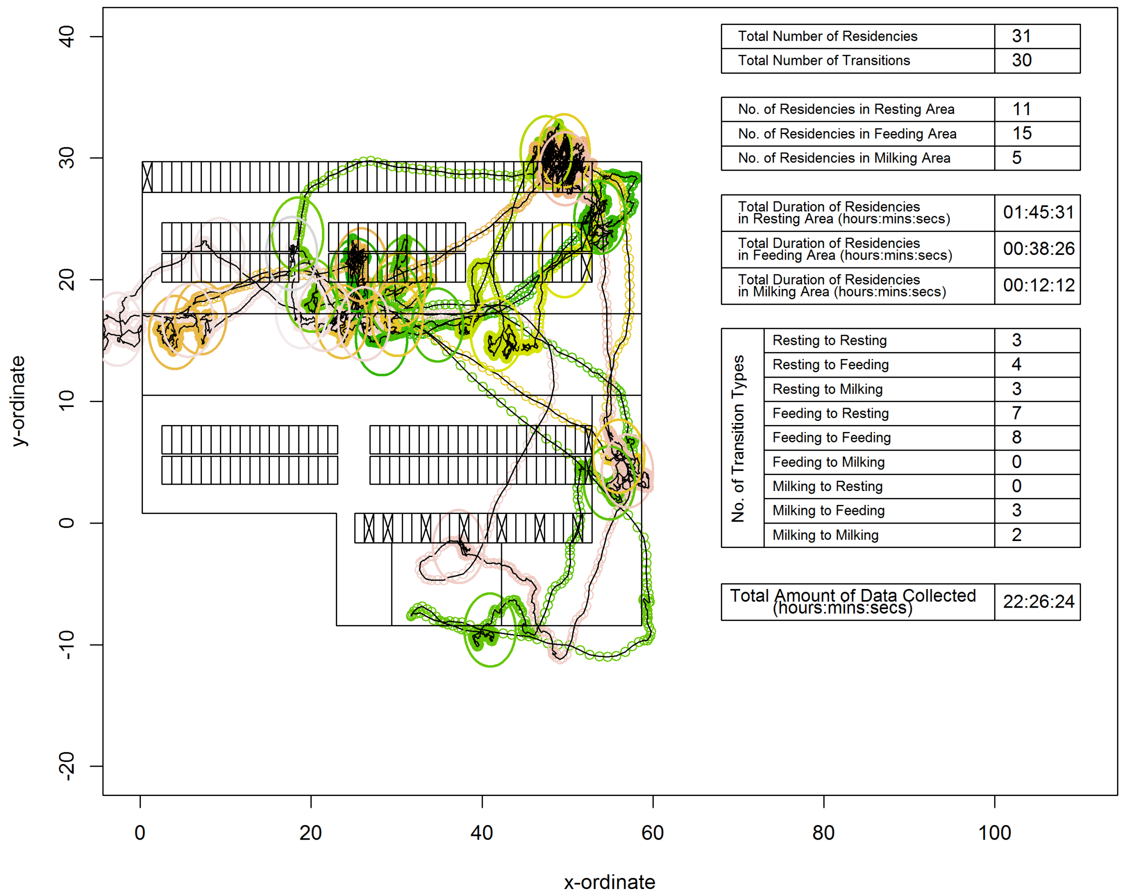
Cow 1891 Day 3 - Out-of-Control Window Size of 11 time points



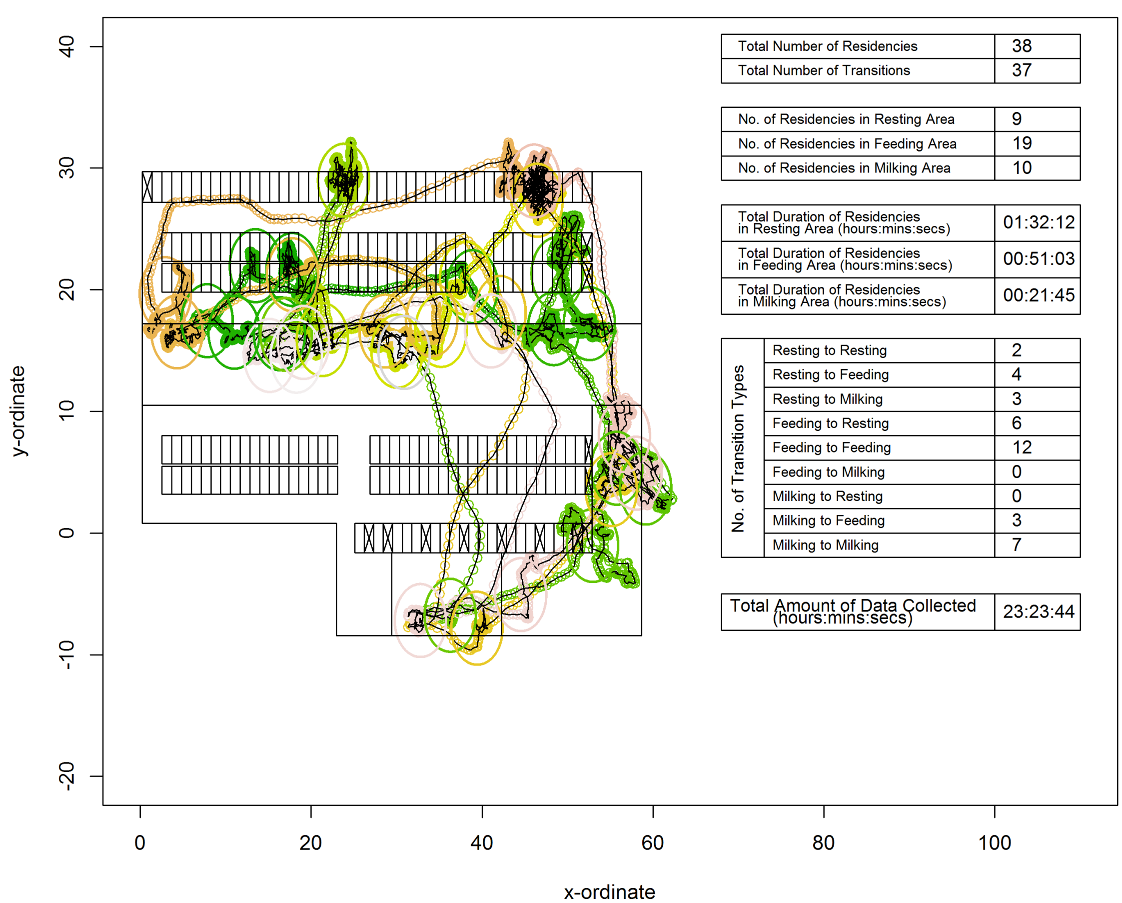
Cow 1891 Day 4 - Out-of-Control Window Size of 11 time points



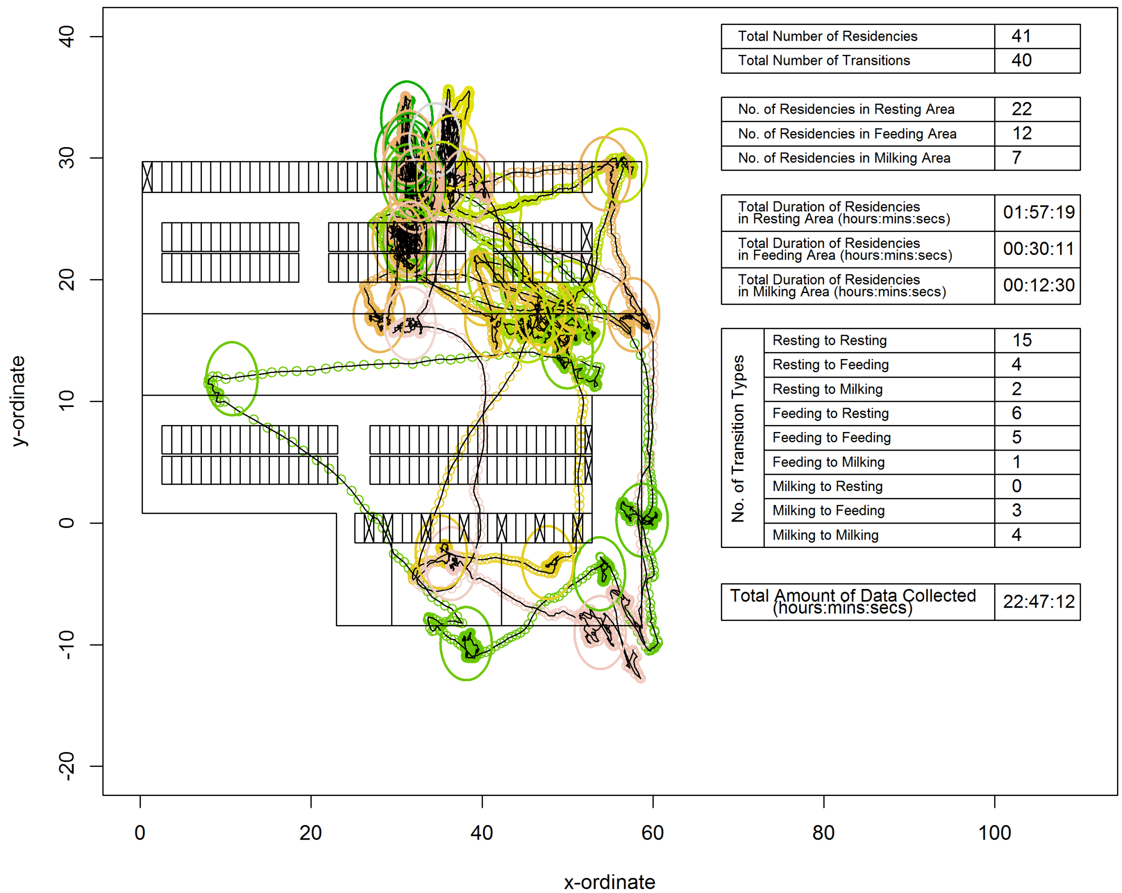
Cow 1891 Day 5 - Out-of-Control Window Size of 11 time points



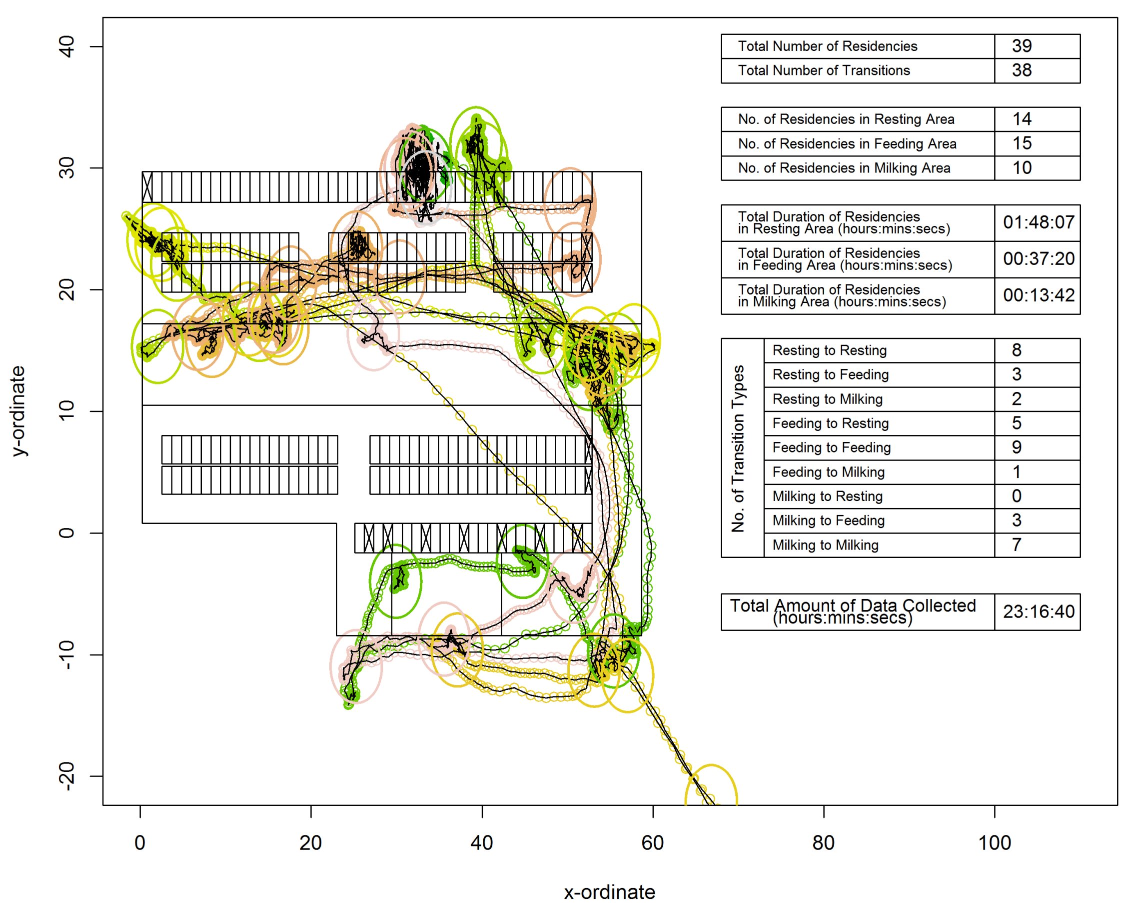
Cow 1891 Day 6 - Out-of-Control Window Size of 11 time points



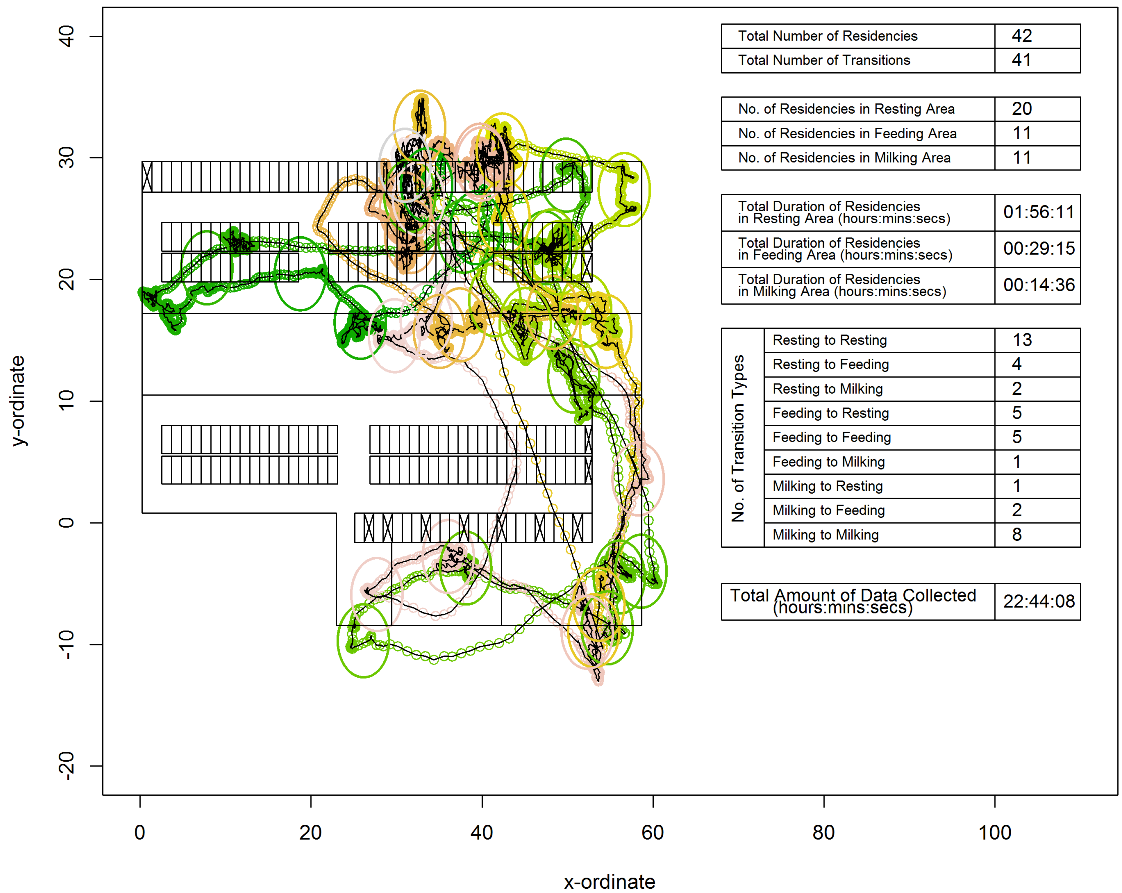
Cow 1078 Day 2 - Out-of-Control Window Size of 11 time points



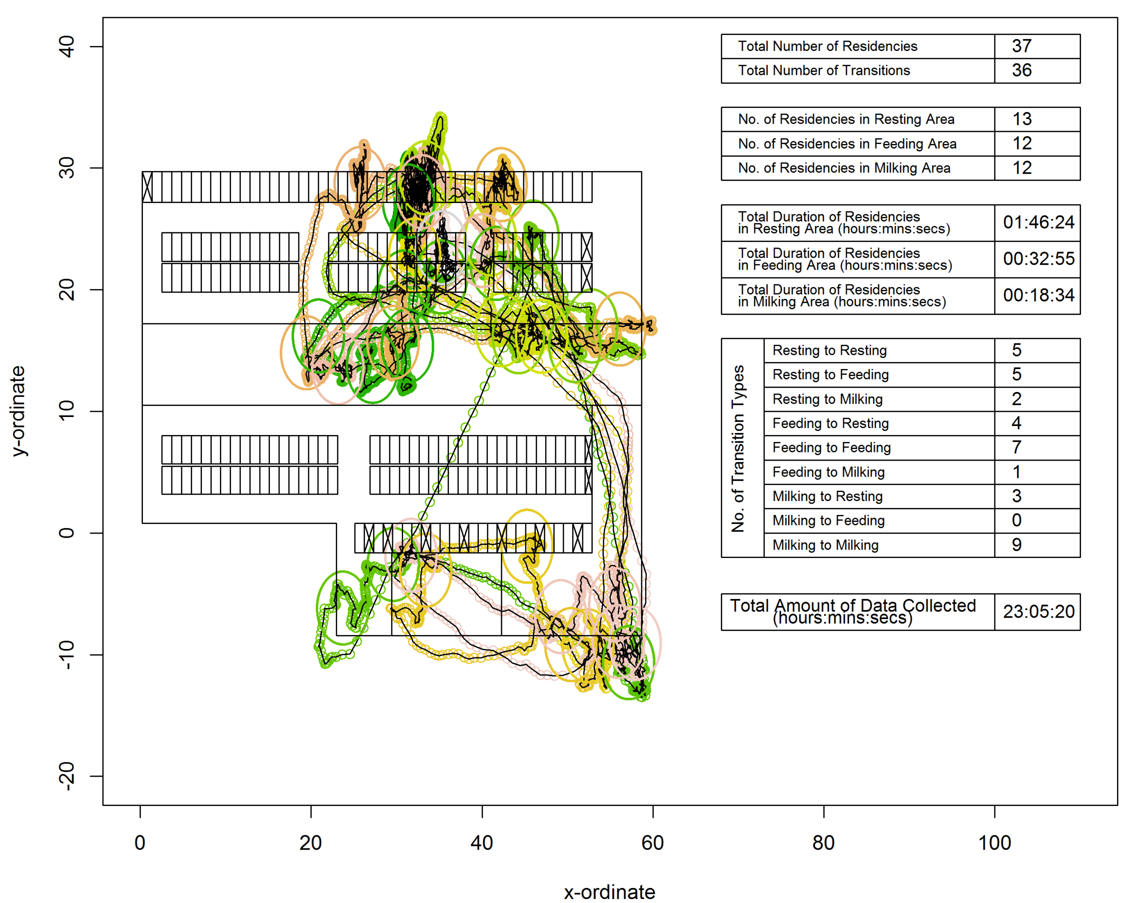
Cow 1078 Day 3 - Out-of-Control Window Size of 11 time points



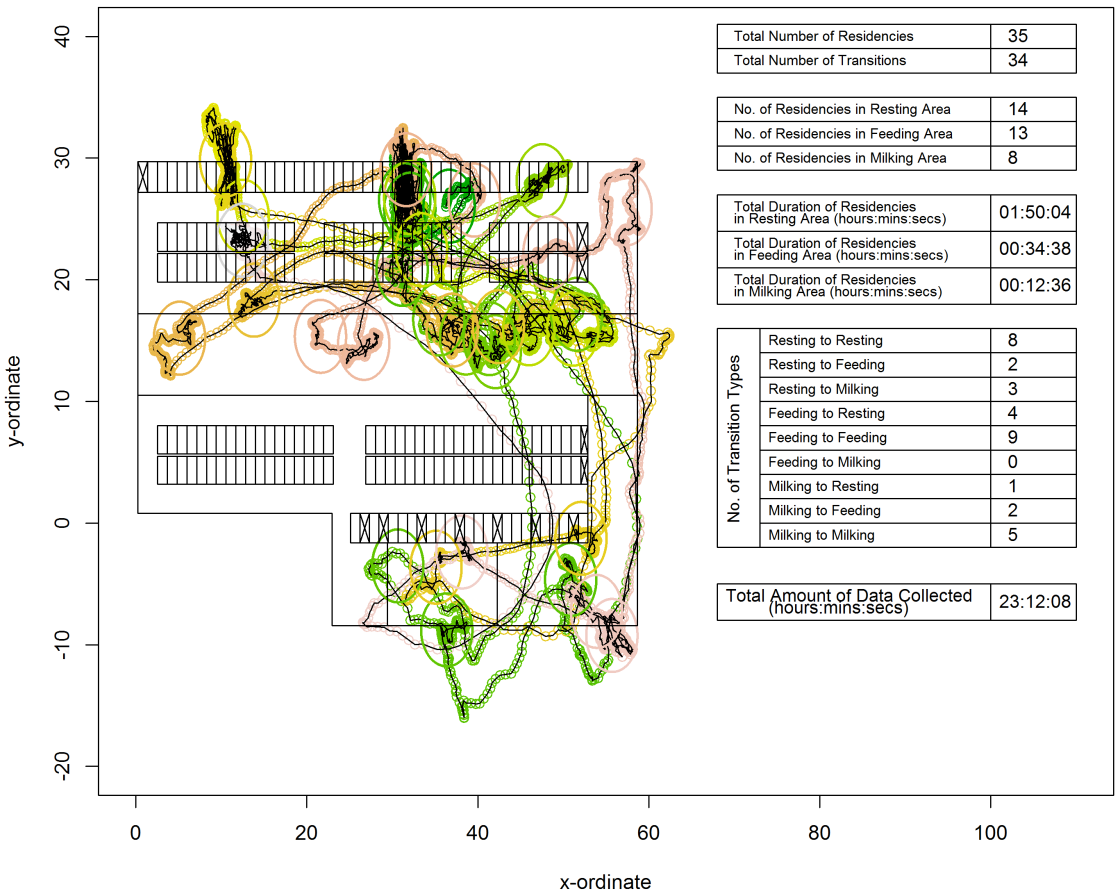
Cow 1078 Day 4 - Out-of-Control Window Size of 11 time points



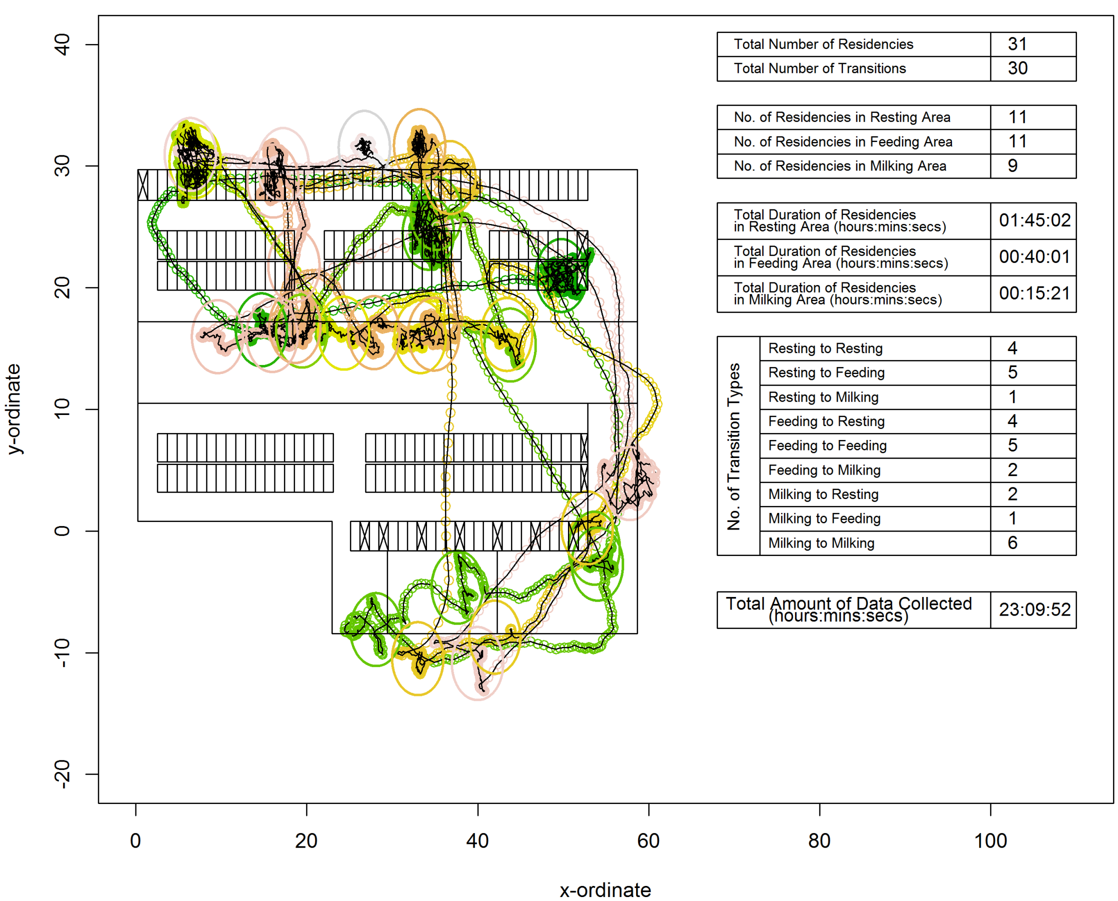
Cow 1078 Day 5 - Out-of-Control Window Size of 11 time points



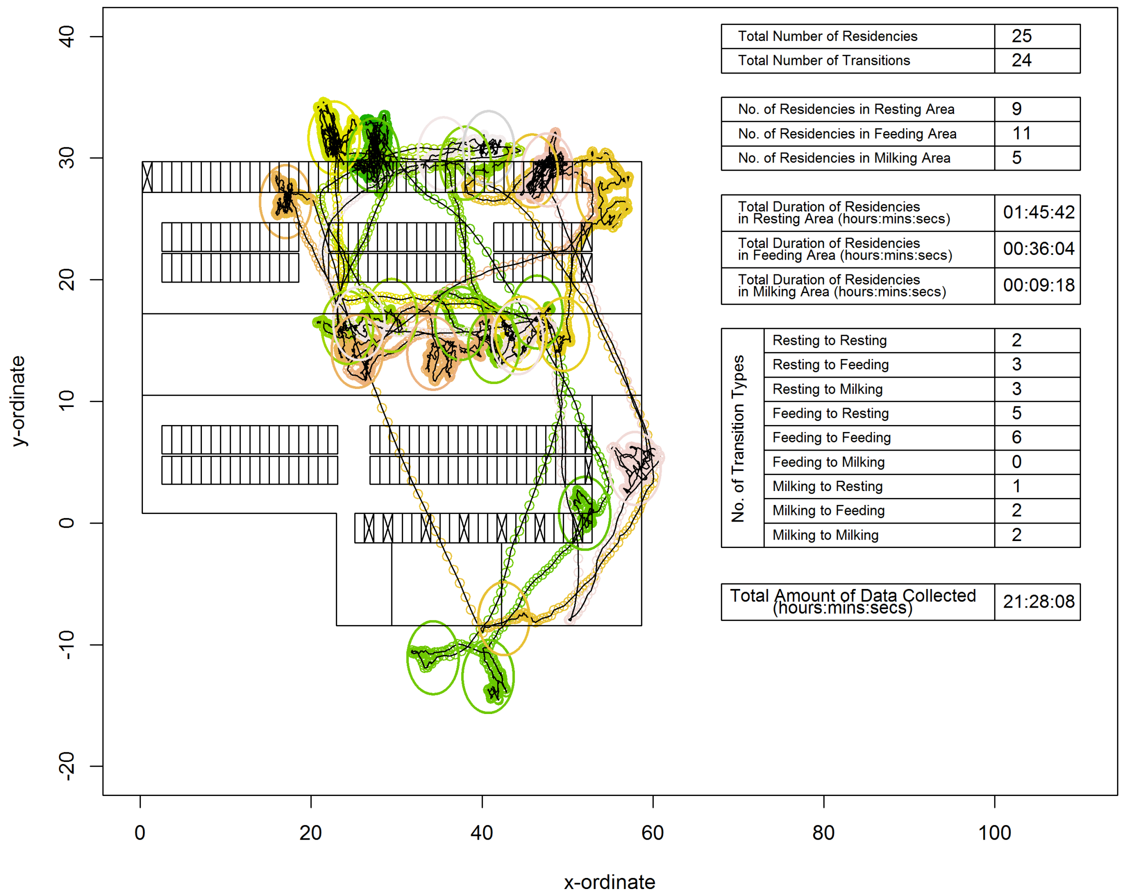
Cow 1078 Day 6 - Out-of-Control Window Size of 11 time points



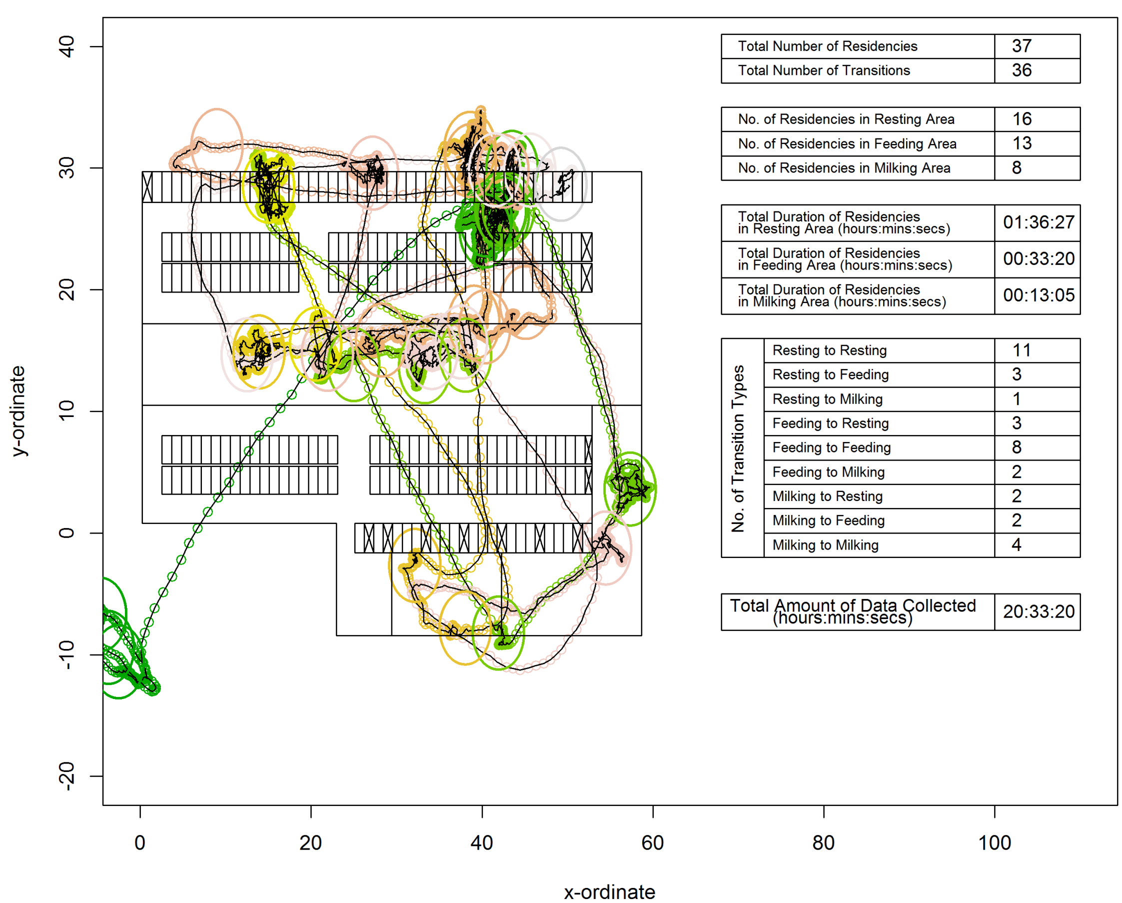
Cow 2616 Day 2 - Out-of-Control Window Size of 11 time points



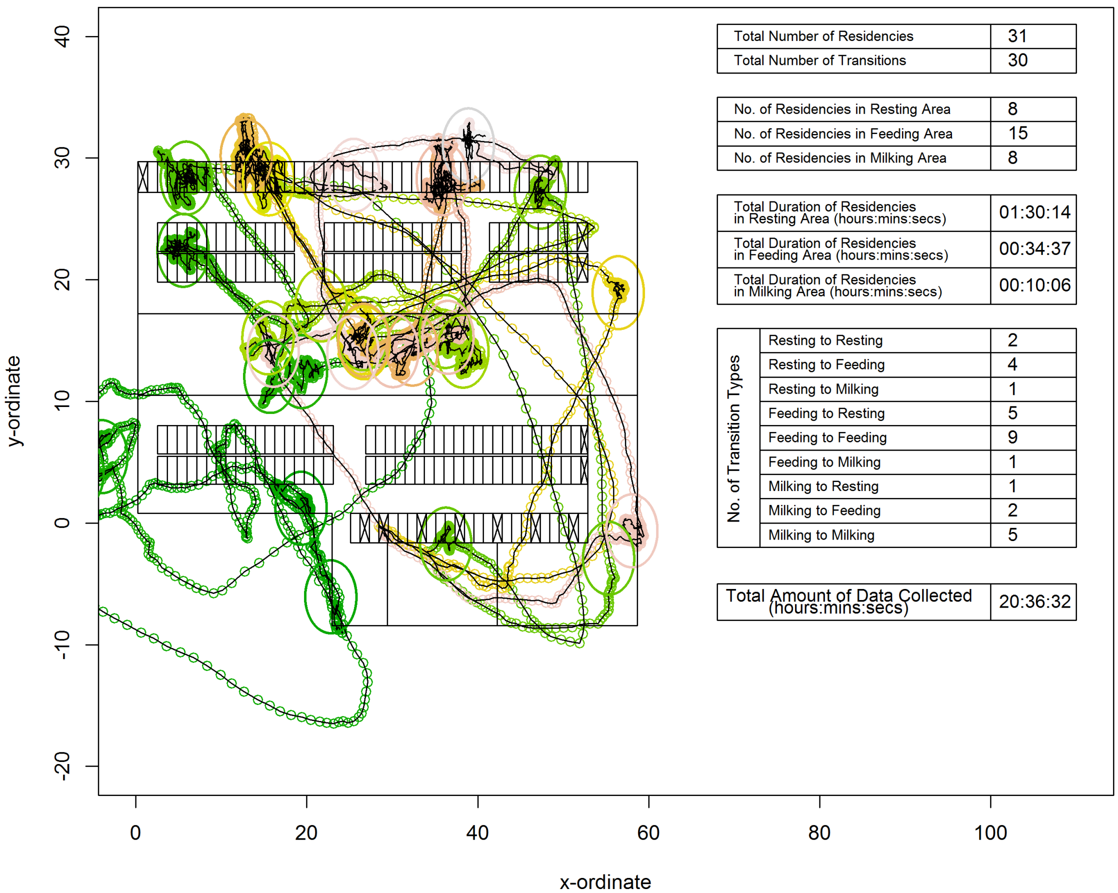
Cow 2616 Day 3 - Out-of-Control Window Size of 11 time points



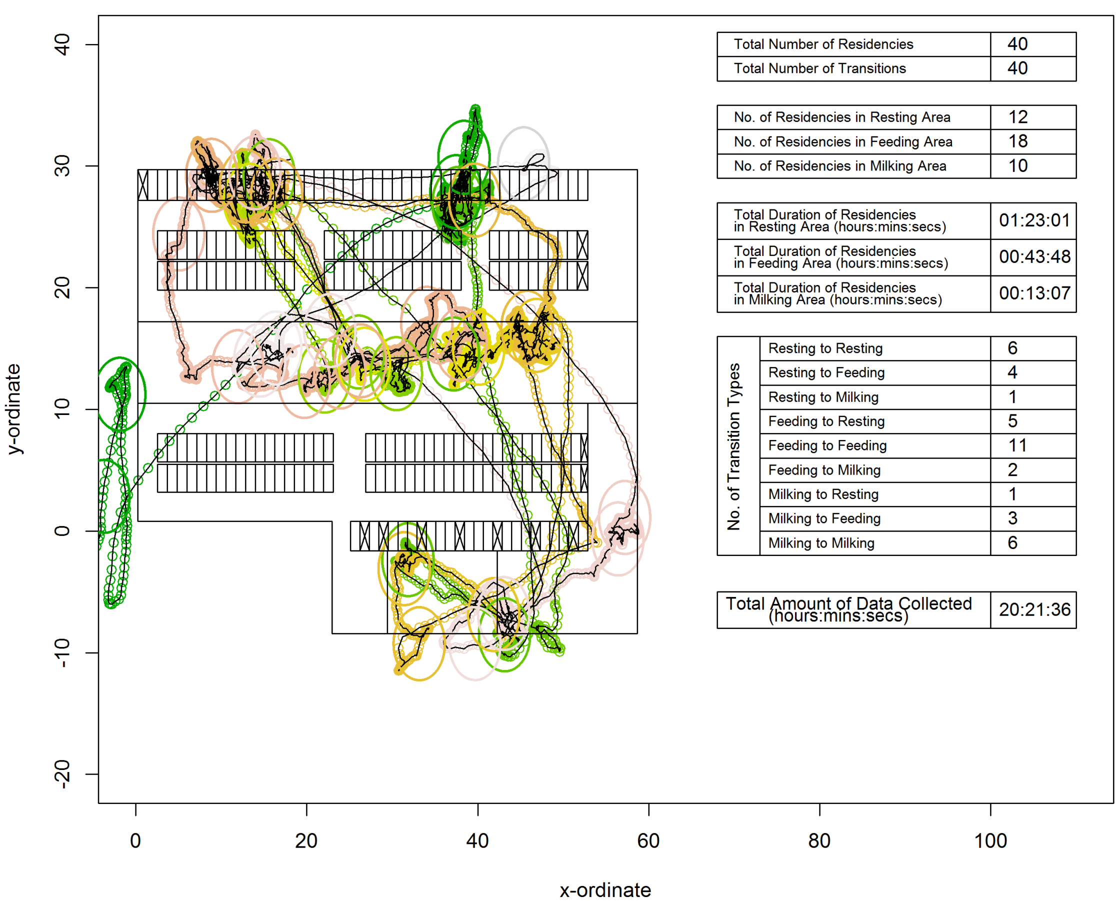
Cow 2616 Day 4 - Out-of-Control Window Size of 11 time points



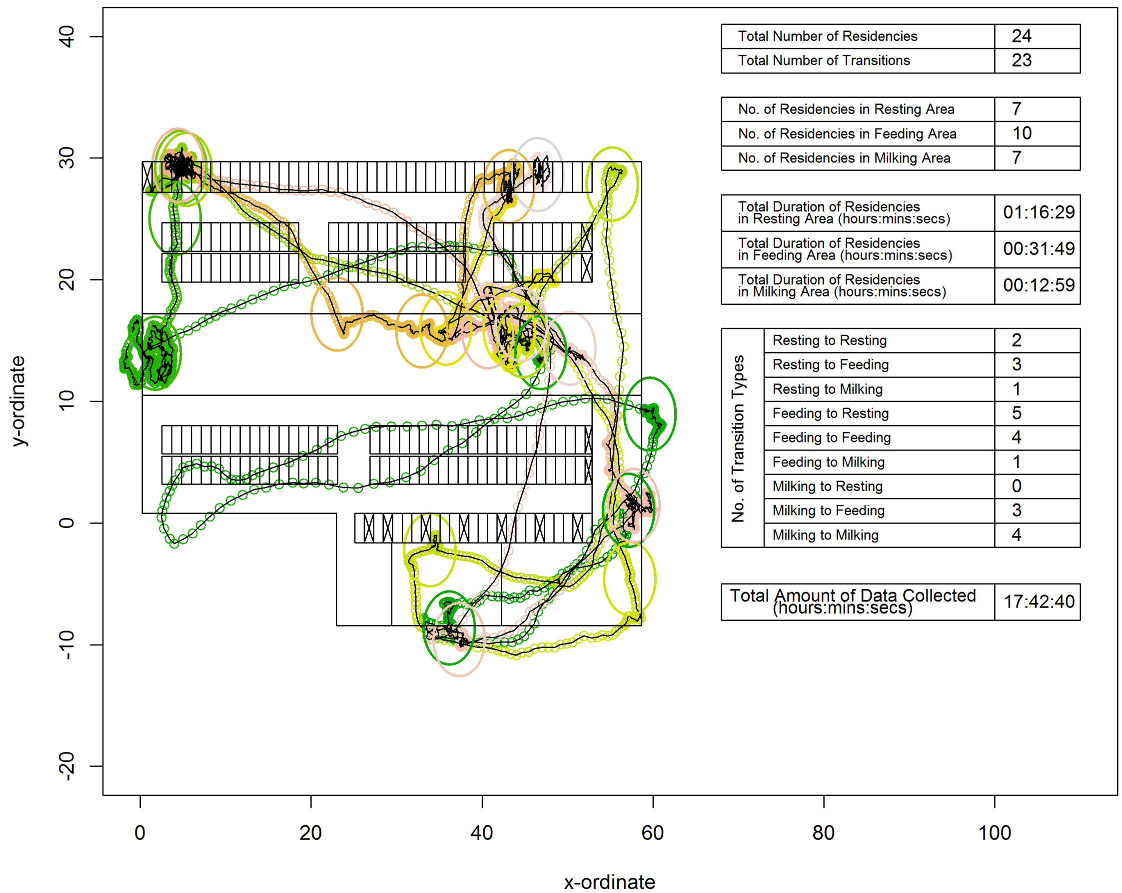
Cow 2616 Day 5 - Out-of-Control Window Size of 11 time points



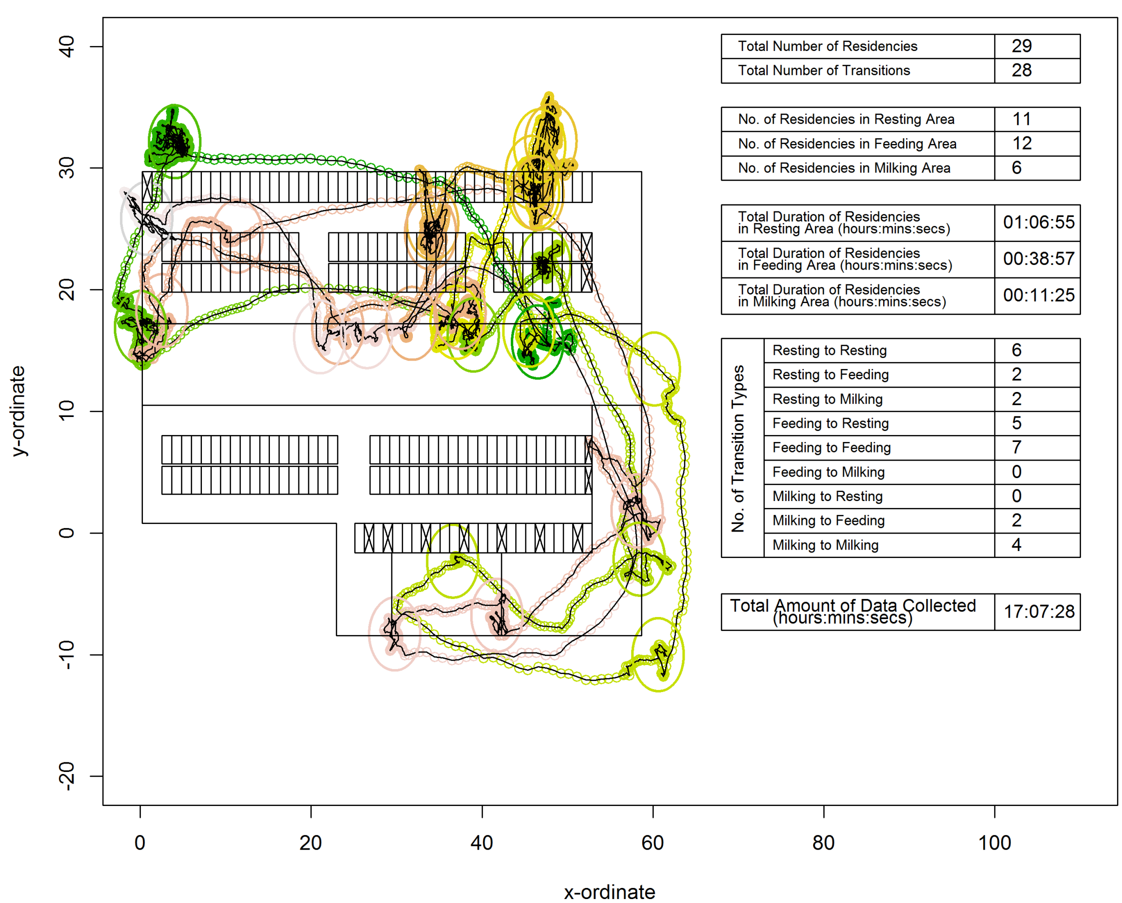
Cow 2616 Day 6 - Out-of-Control Window Size of 11 time points



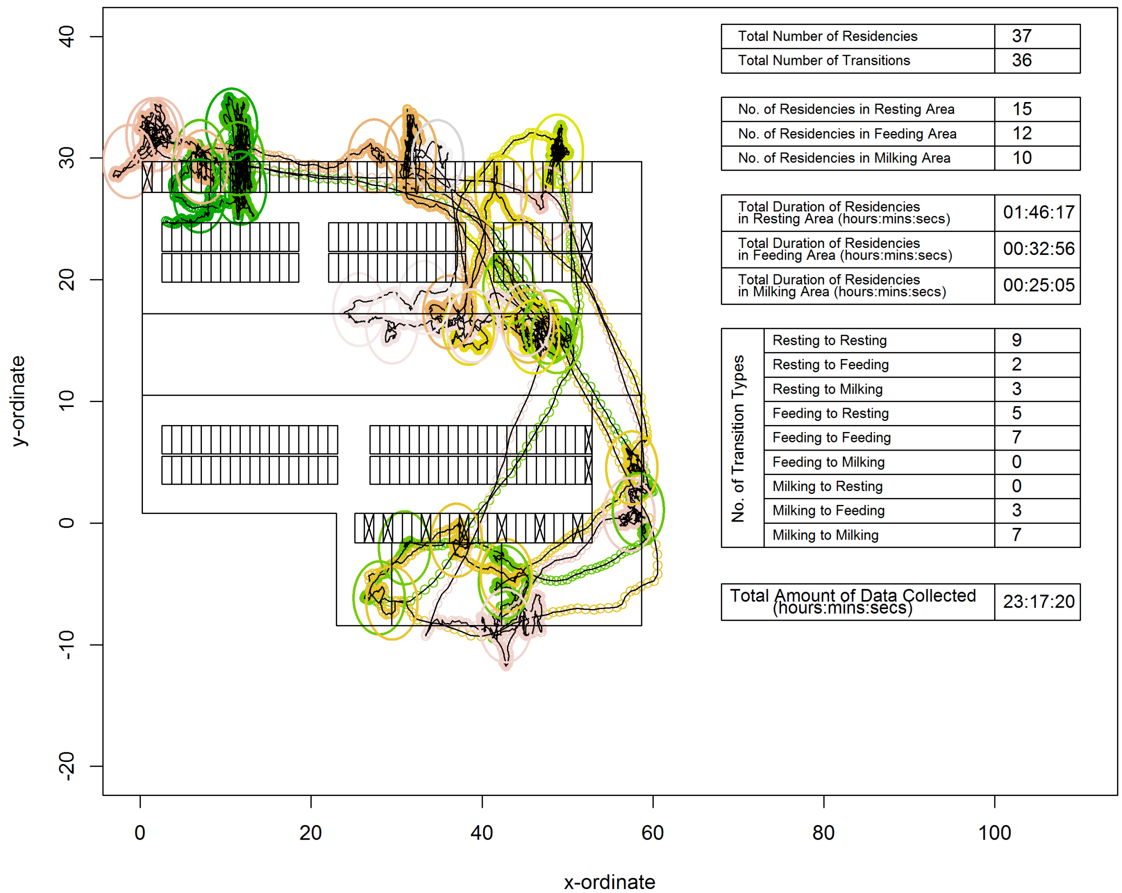
Cow 2302 Day 2 - Out-of-Control Window Size of 11 time points



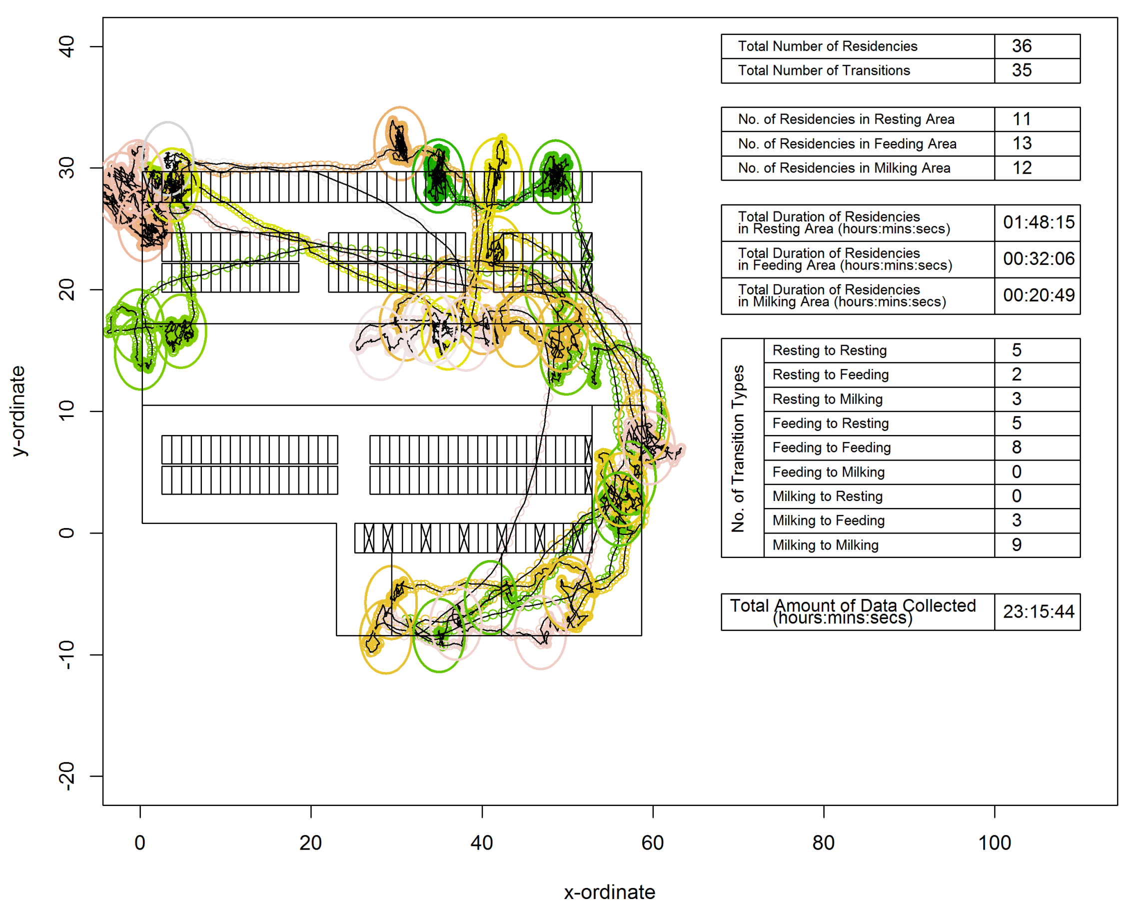
Cow 2302 Day 3 - Out-of-Control Window Size of 11 time points

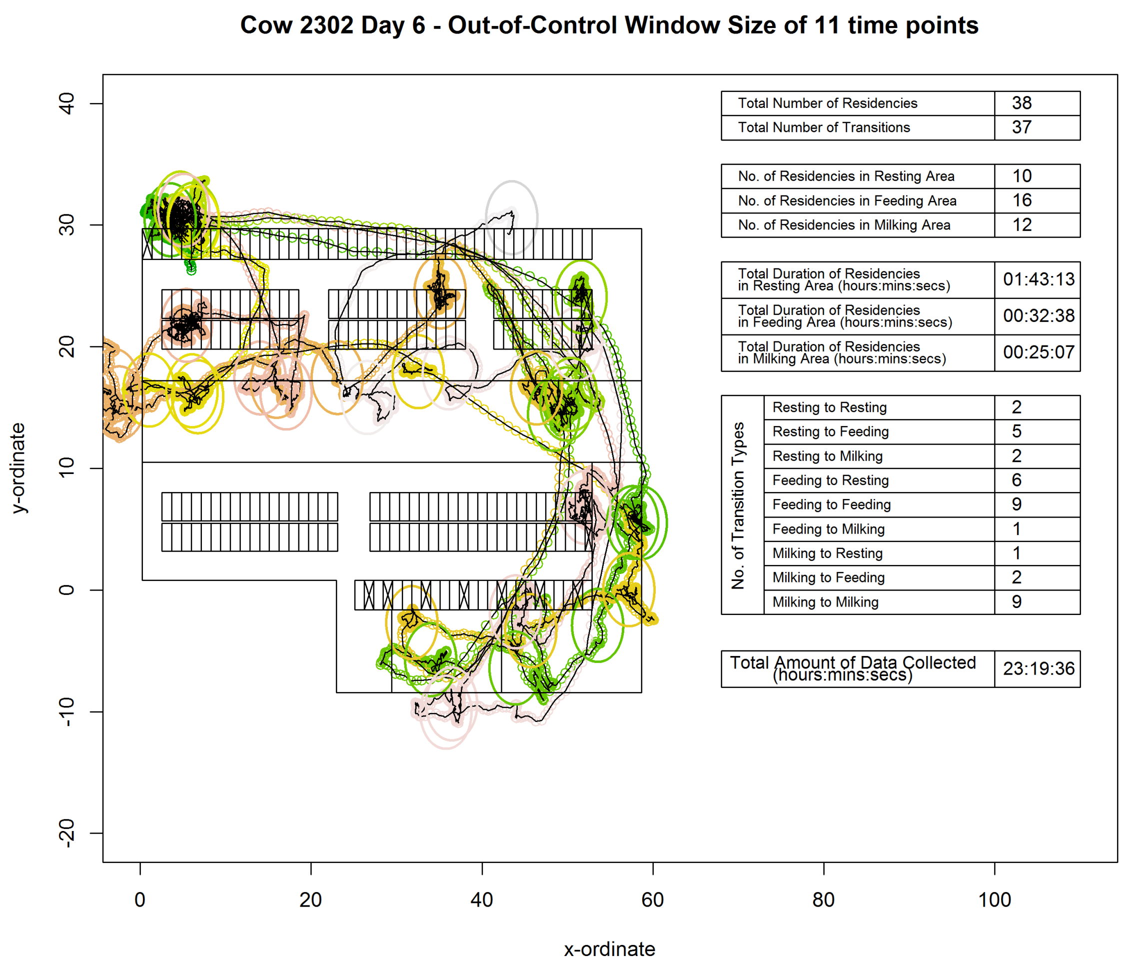


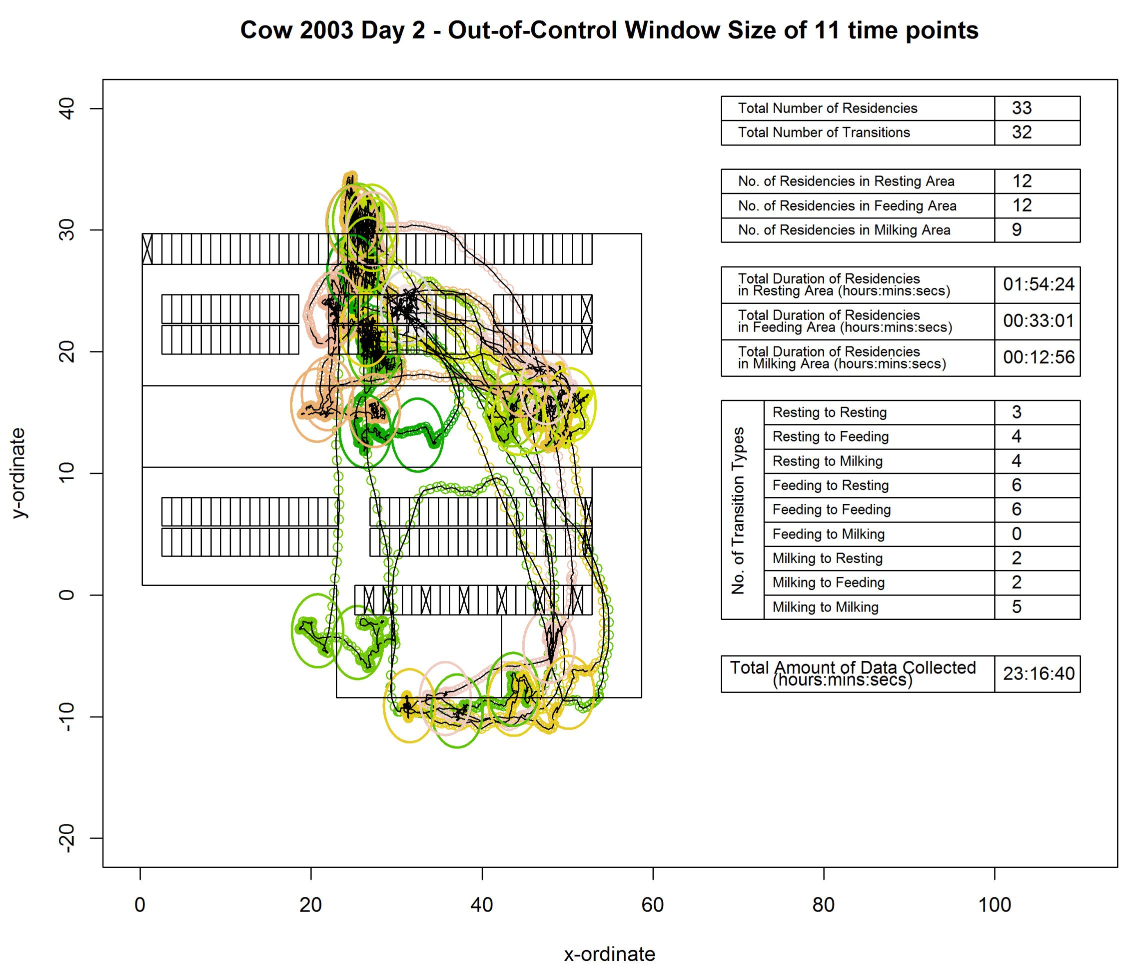
Cow 2302 Day 4 - Out-of-Control Window Size of 11 time points



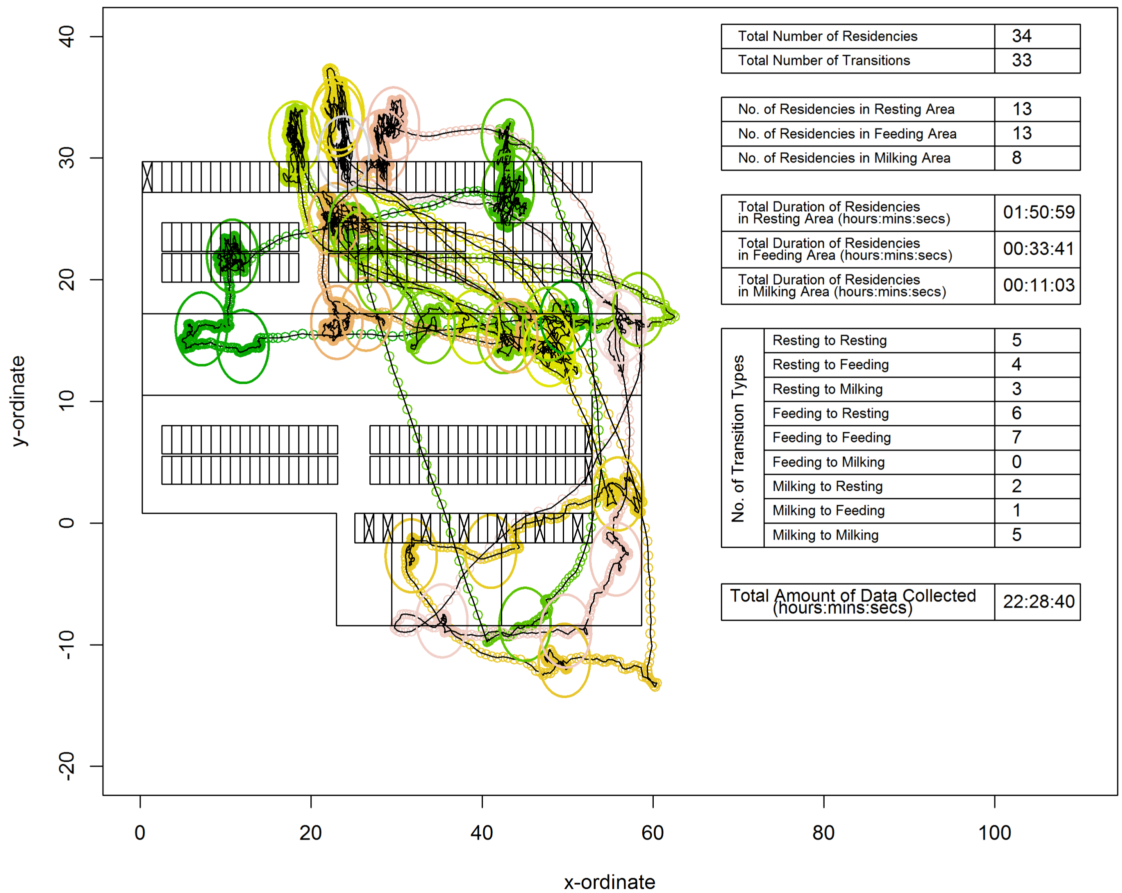
Cow 2302 Day 5 - Out-of-Control Window Size of 11 time points



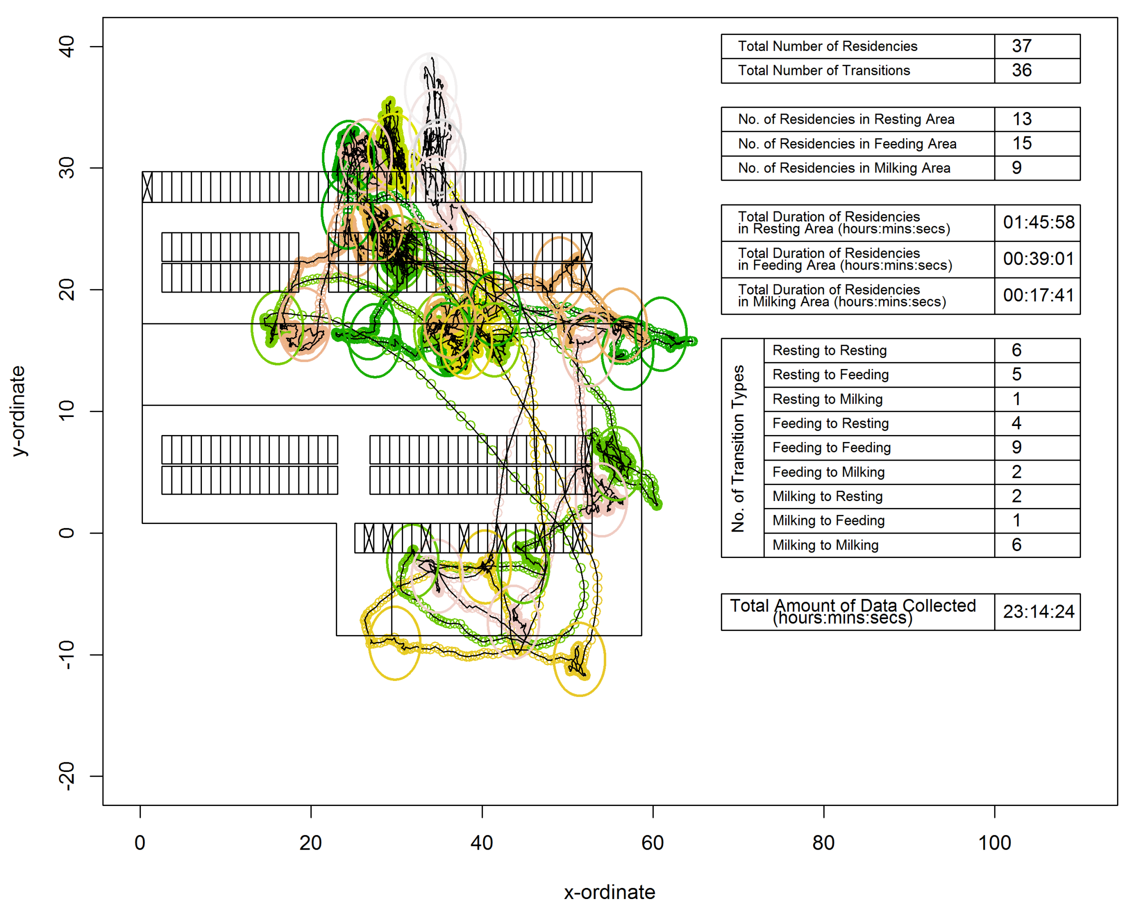

Cow 2302 Day 6 - Out-of-Control Window Size of 11 time points
Total Number of Residencies 38
Total Number of Transitions 37
No. of Residencies in Resting Area 10
No. of Residencies in Feeding Area 16
No. of Residencies in Milking Area 12
Total Duration of Residencies in Resting Area (hours:mins:secs) 01:43:13
Total Duration of Residencies in Feeding Area (hours:mins:secs) 00:32:38
Total Duration of Residencies in Milking Area (hours:mins:secs) 00:25:07
No. of Transition Types
Resting to Resting 2
Resting to Feeding 5
Resting to Milking 2
Feeding to Resting 6
Feeding to Feeding 9
Feeding to Milking 1
Milking to Resting 1
Milking to Feeding 2
Milking to Milking 9
Total Amount of Data Collected (hours:mins:secs) 23:19:36
y-ordinate
x-ordinate
40
30
20
10
0
-10
-20
0
20
40
60
80
100


Cow 2003 Day 2 - Out-of-Control Window Size of 11 time points
Total Number of Residencies 33
Total Number of Transitions 32
No. of Residencies in Resting Area 12
No. of Residencies in Feeding Area 12
No. of Residencies in Milking Area 9
Total Duration of Residencies in Resting Area (hours:mins:secs) 01:54:24
Total Duration of Residencies in Feeding Area (hours:mins:secs) 00:33:01
Total Duration of Residencies in Milking Area (hours:mins:secs) 00:12:56
No. of Transition Types
Resting to Resting 3
Resting to Feeding 4
Resting to Milking 4
Feeding to Resting 6
Feeding to Feeding 6
Feeding to Milking 0
Milking to Resting 2
Milking to Feeding 2
Milking to Milking 5
Total Amount of Data Collected (hours:mins:secs) 23:16:40
y-ordinate
x-ordinate
40
30
20
10
0
-10
-20
0
20
40
60
80
100

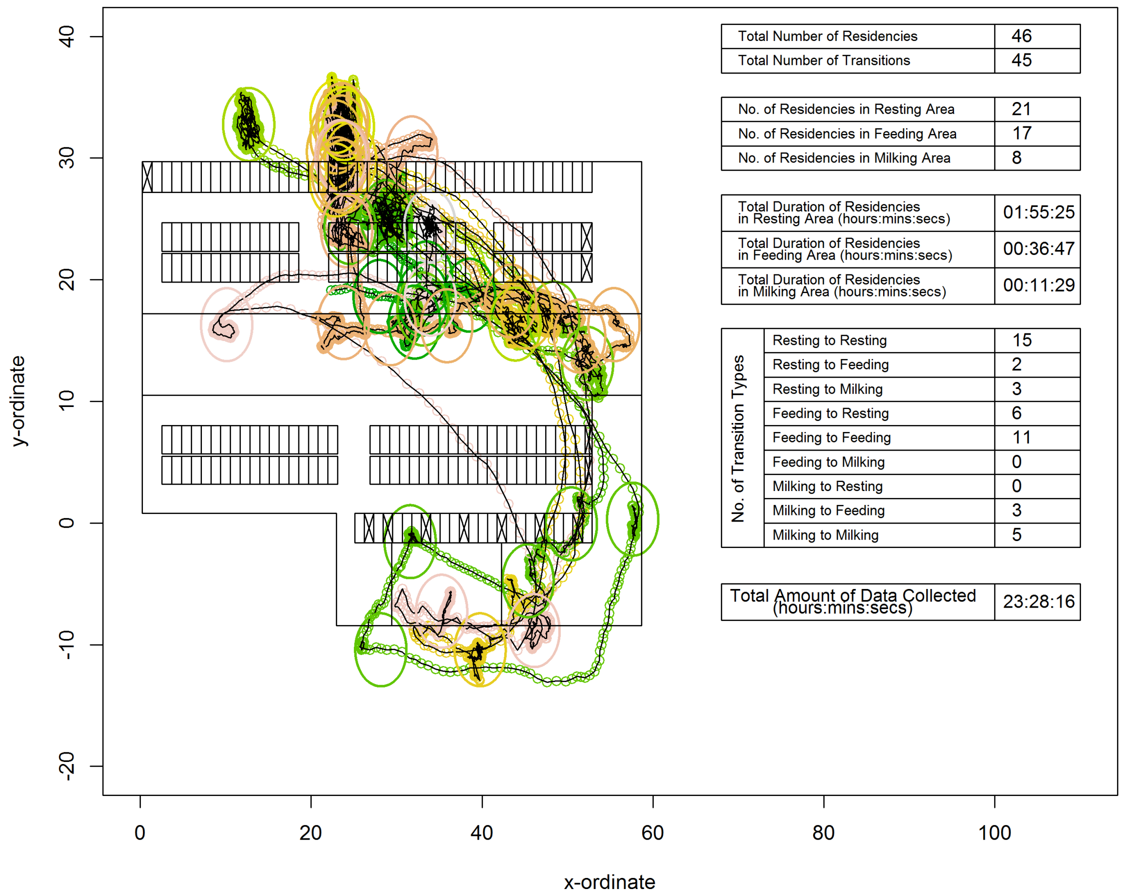
Cow 2003 Day 3 - Out-of-Control Window Size of 11 time points



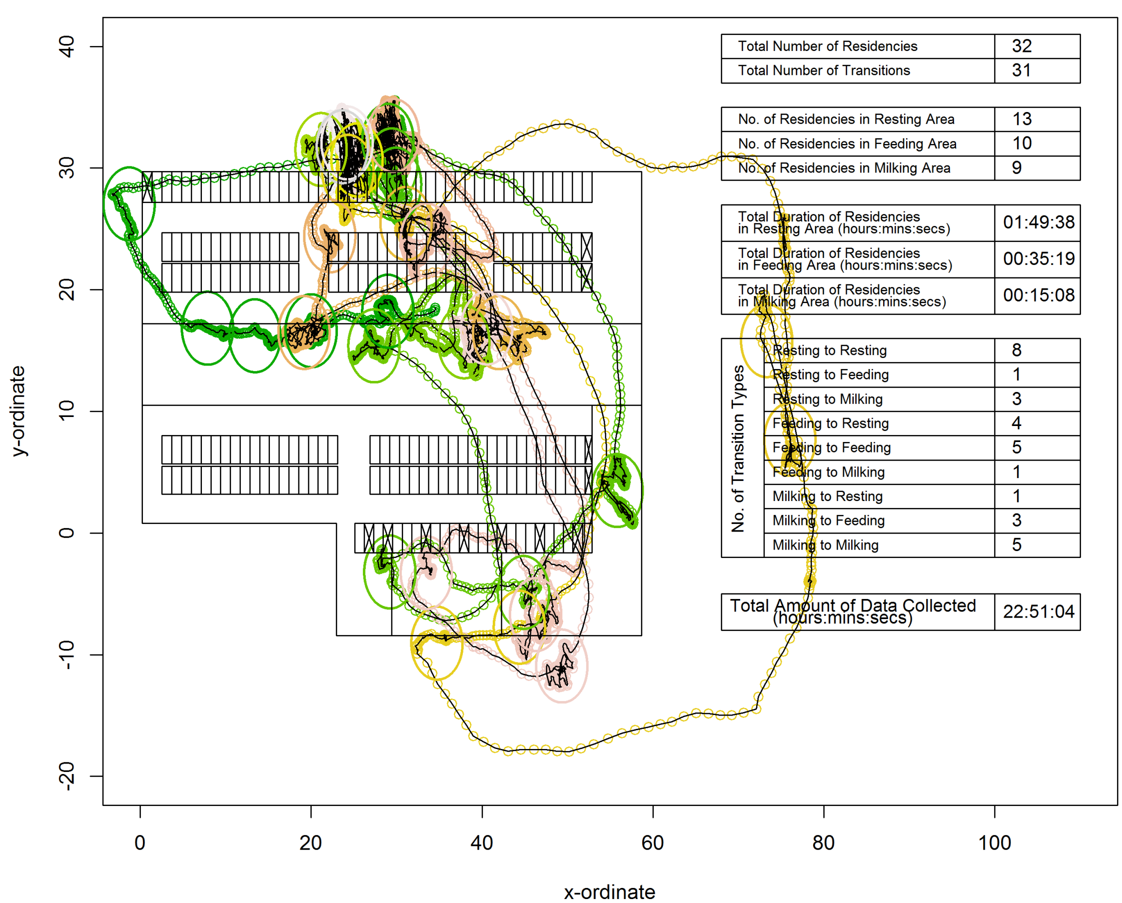
Cow 2003 Day 4 - Out-of-Control Window Size of 11 time points



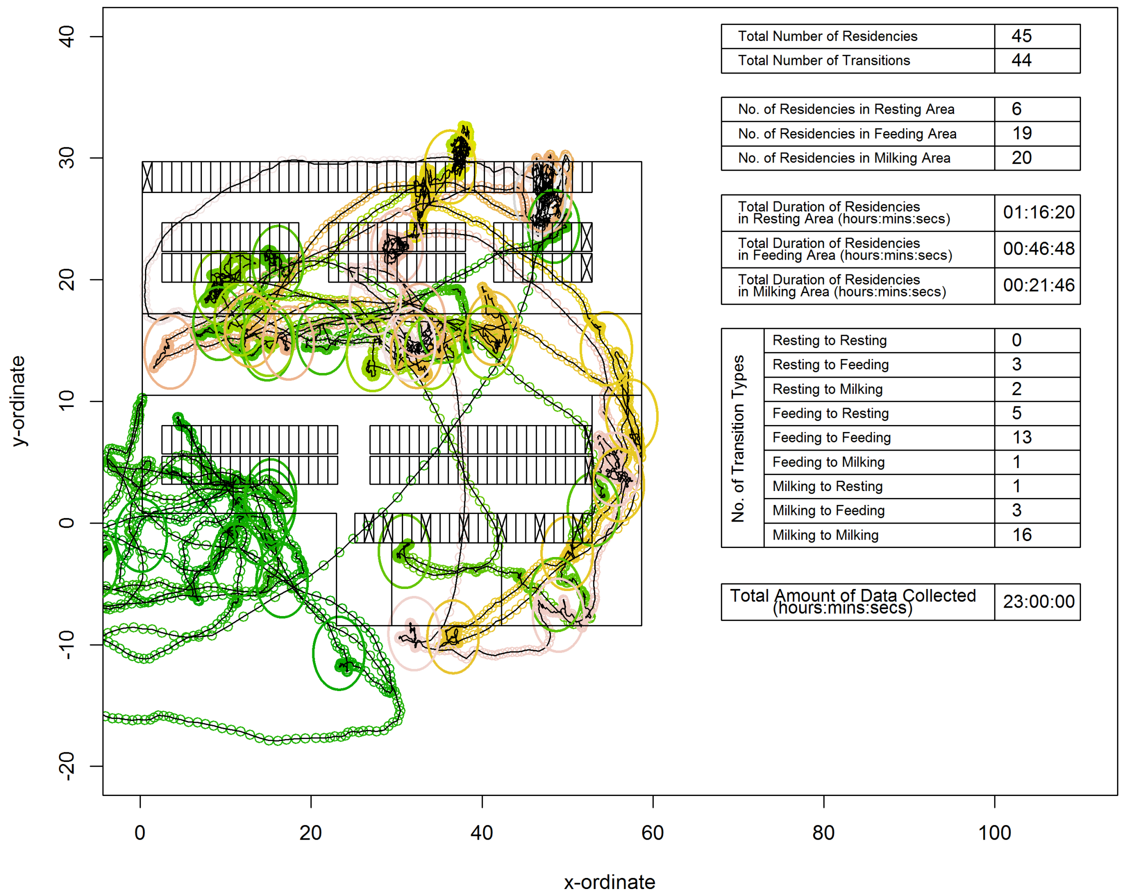
Cow 2003 Day 5 - Out-of-Control Window Size of 11 time points



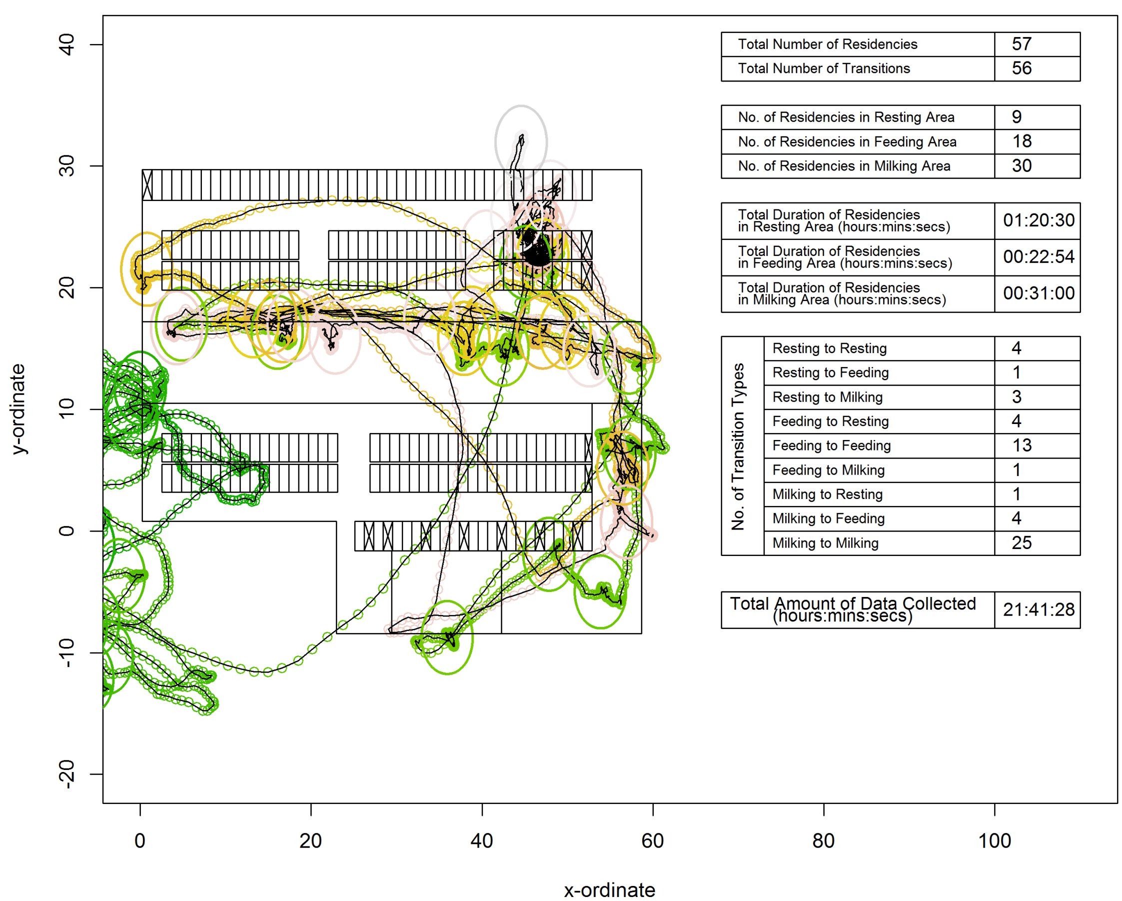
Cow 2003 Day 6 - Out-of-Control Window Size of 11 time points



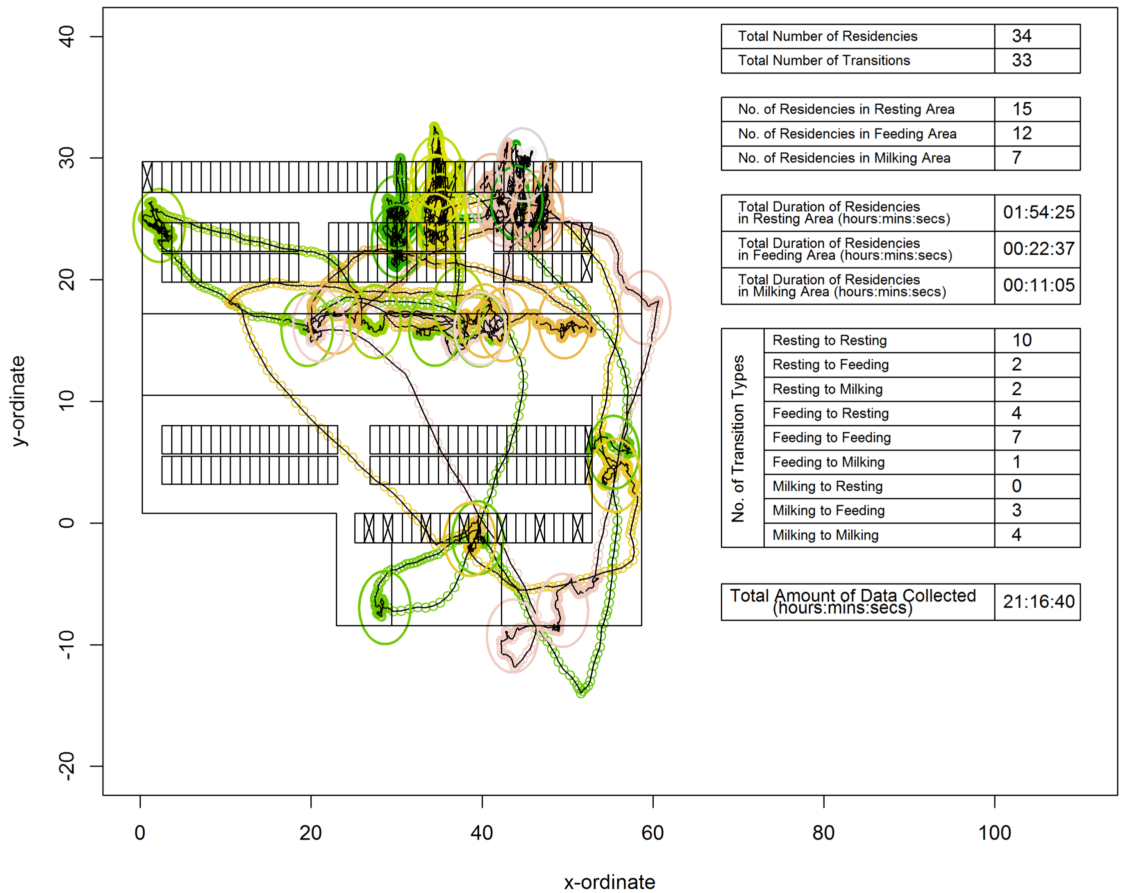
Cow 1184 Day 2 - Out-of-Control Window Size of 11 time points



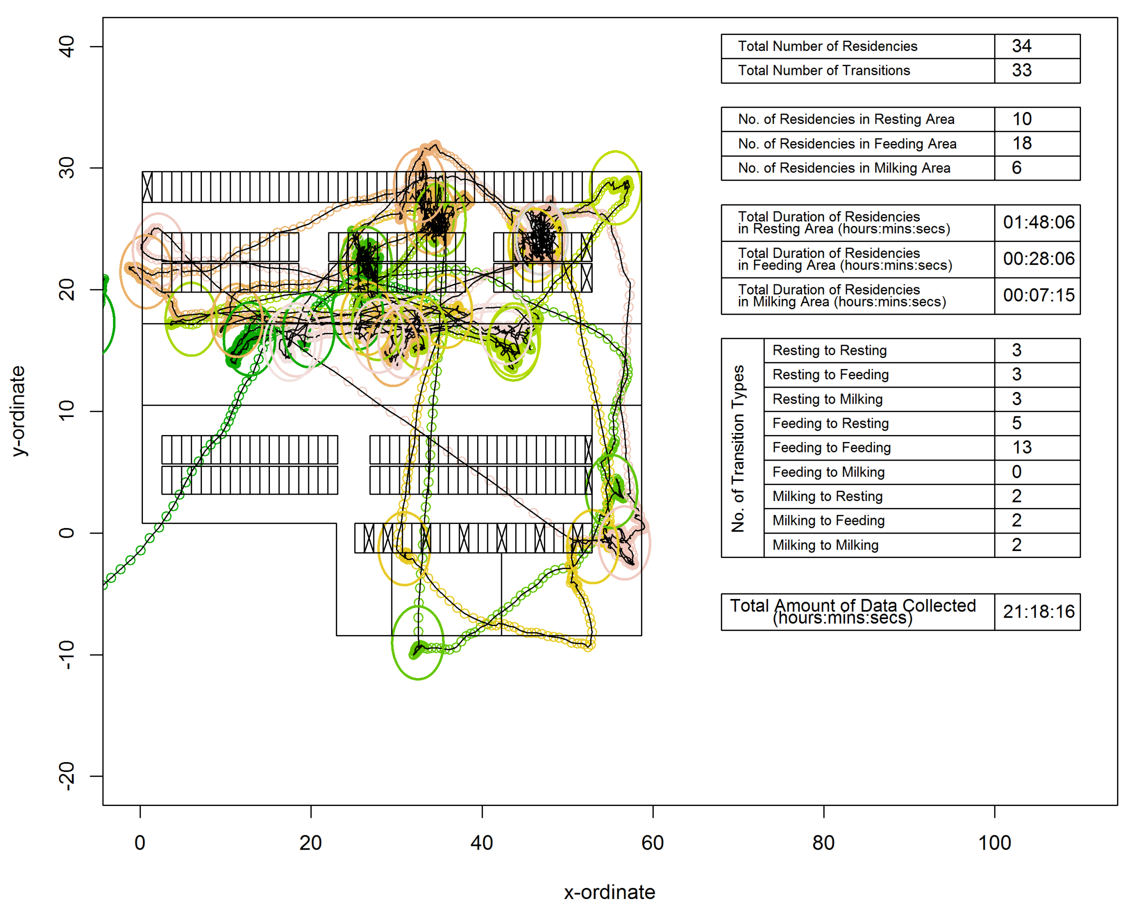
Cow 1184 Day 3 - Out-of-Control Window Size of 11 time points



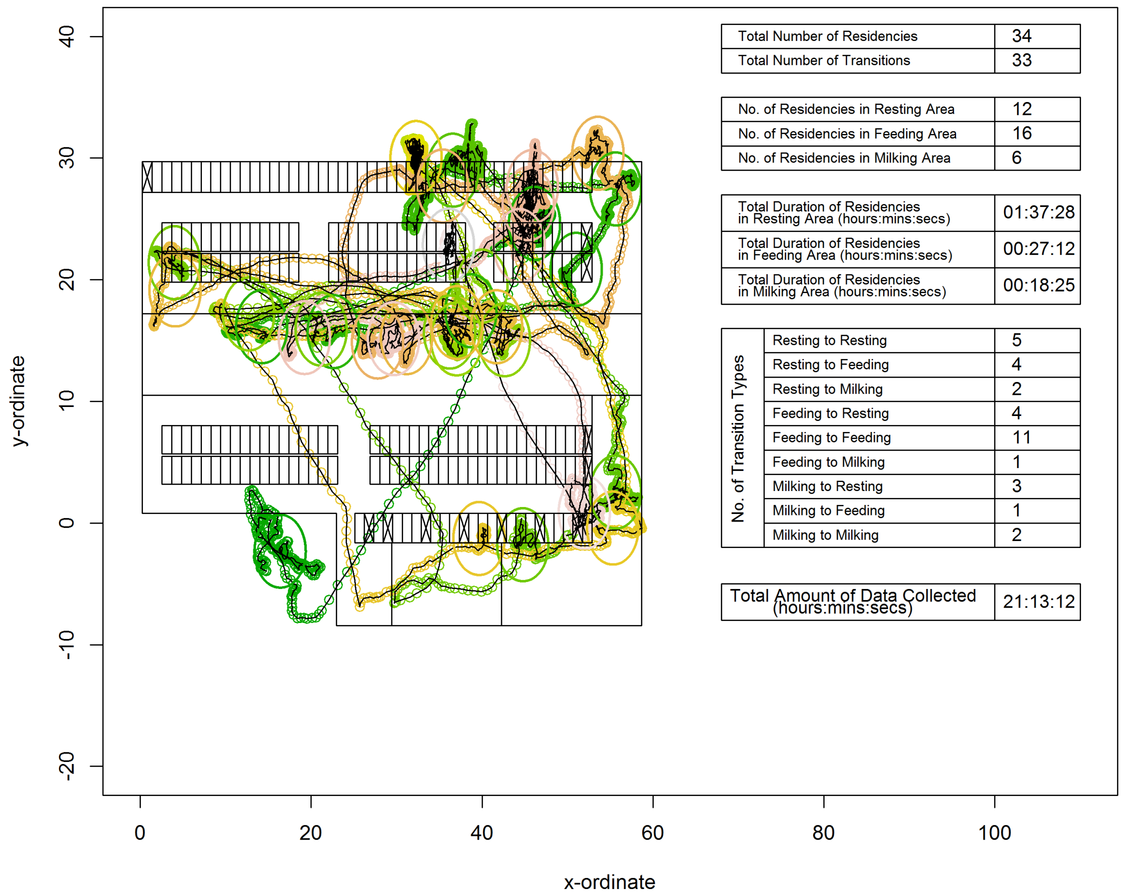
Cow 1184 Day 4 - Out-of-Control Window Size of 11 time points



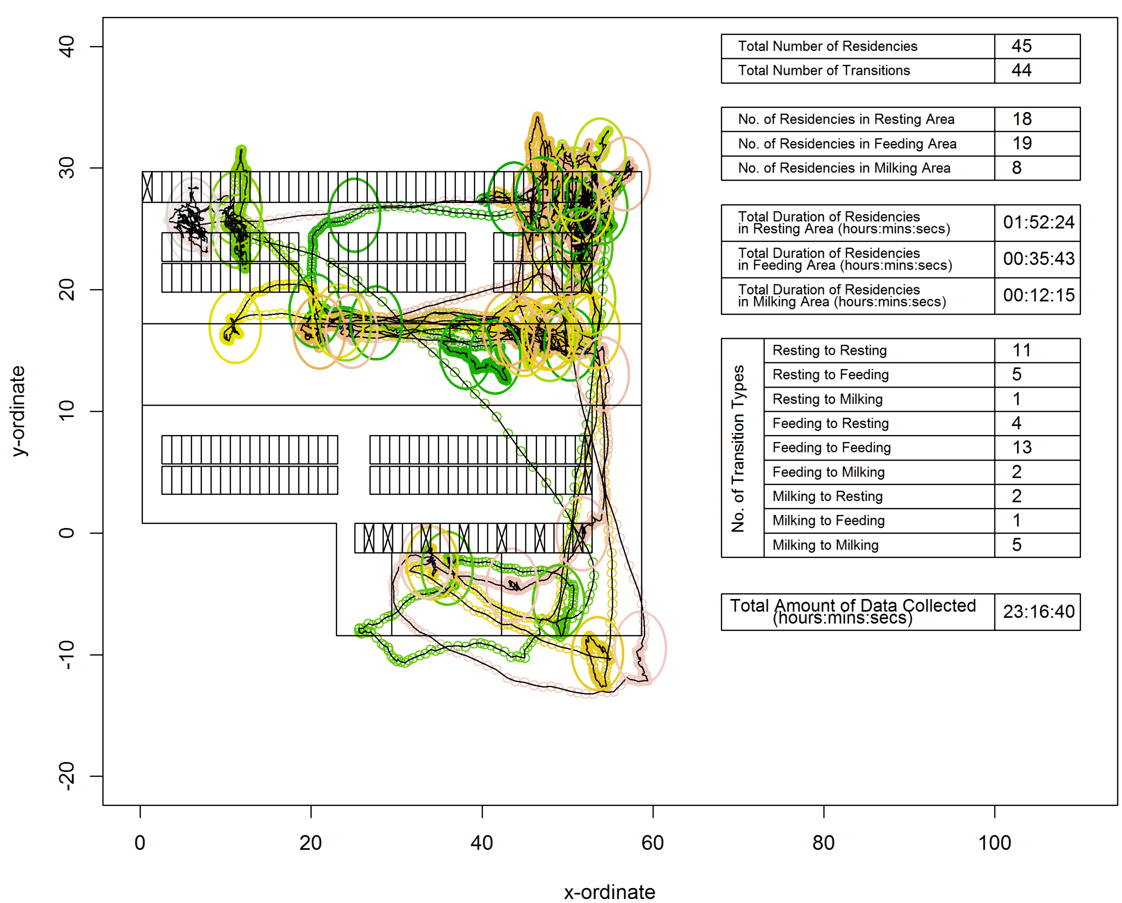
Cow 1184 Day 5 - Out-of-Control Window Size of 11 time points



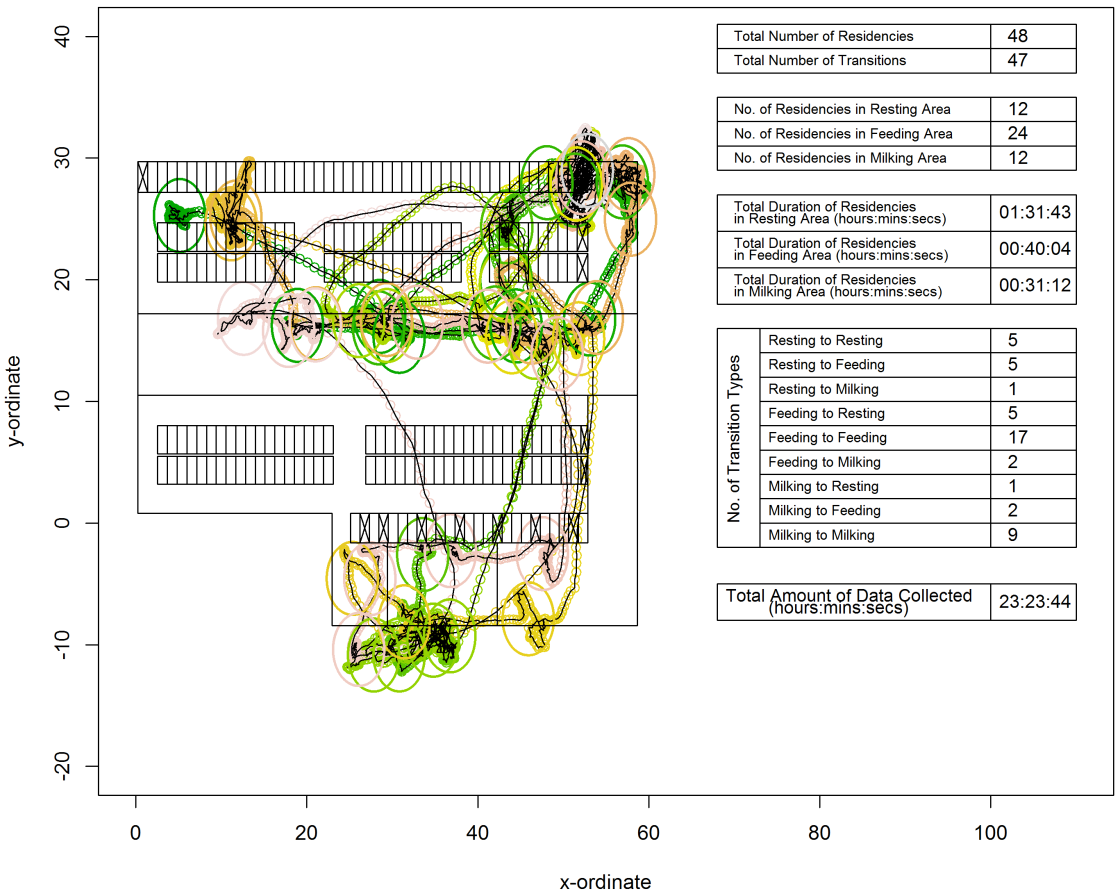
Cow 1184 Day 6 - Out-of-Control Window Size of 11 time points



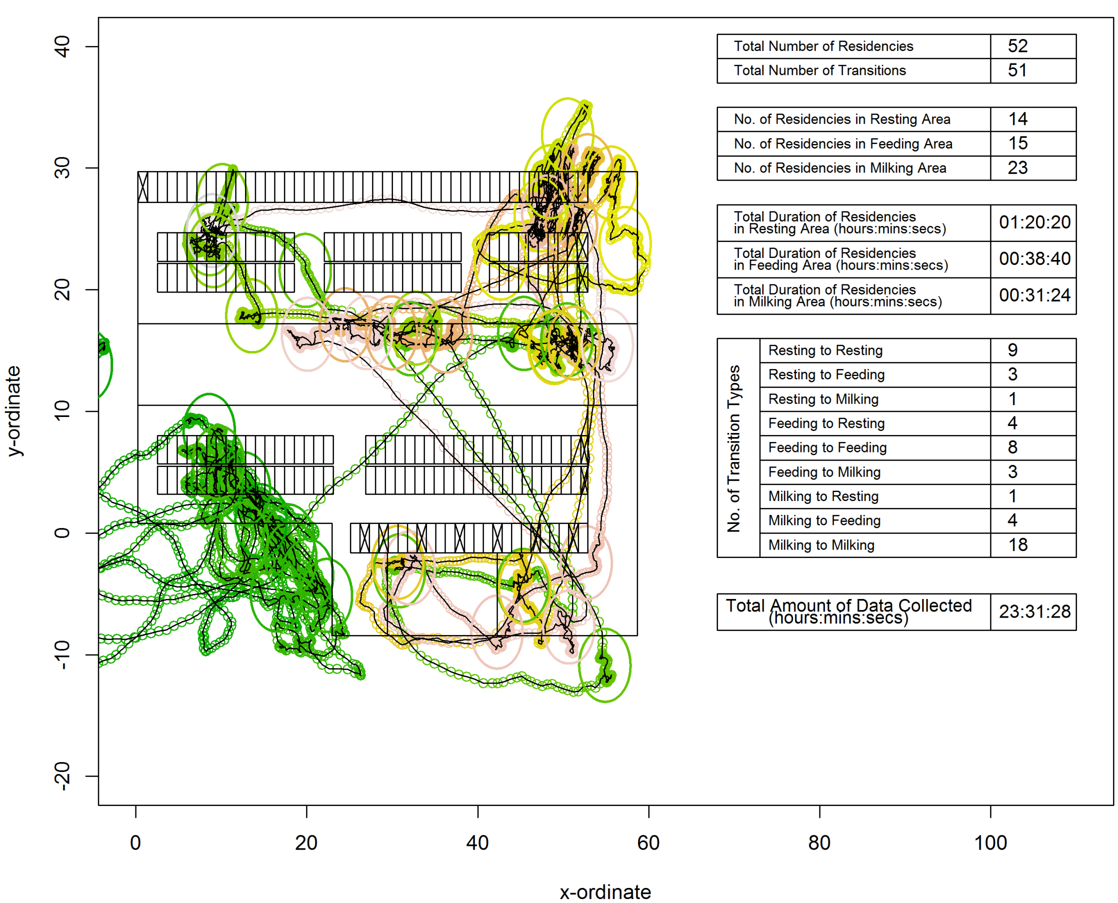
Cow 1340 Day 2 - Out-of-Control Window Size of 11 time points



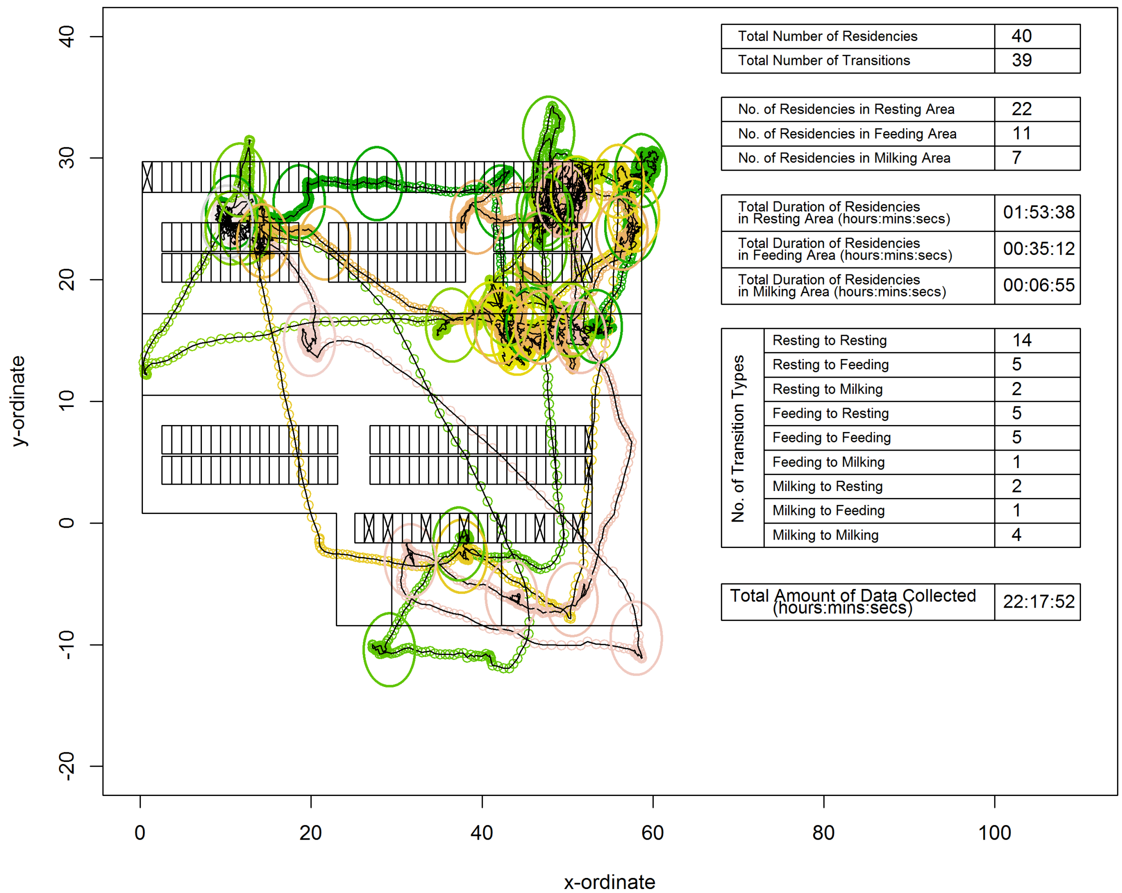
Cow 1340 Day 3 - Out-of-Control Window Size of 11 time points



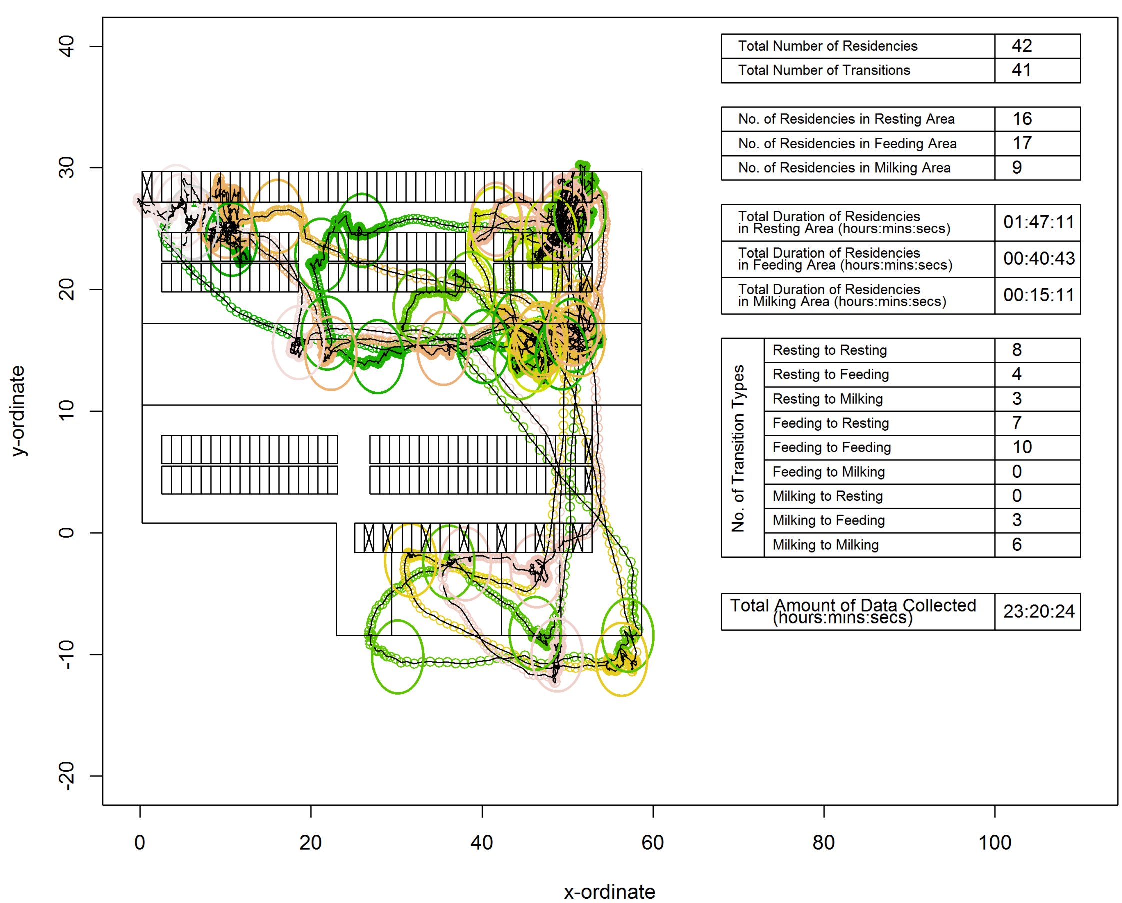
Cow 1340 Day 4 - Out-of-Control Window Size of 11 time points



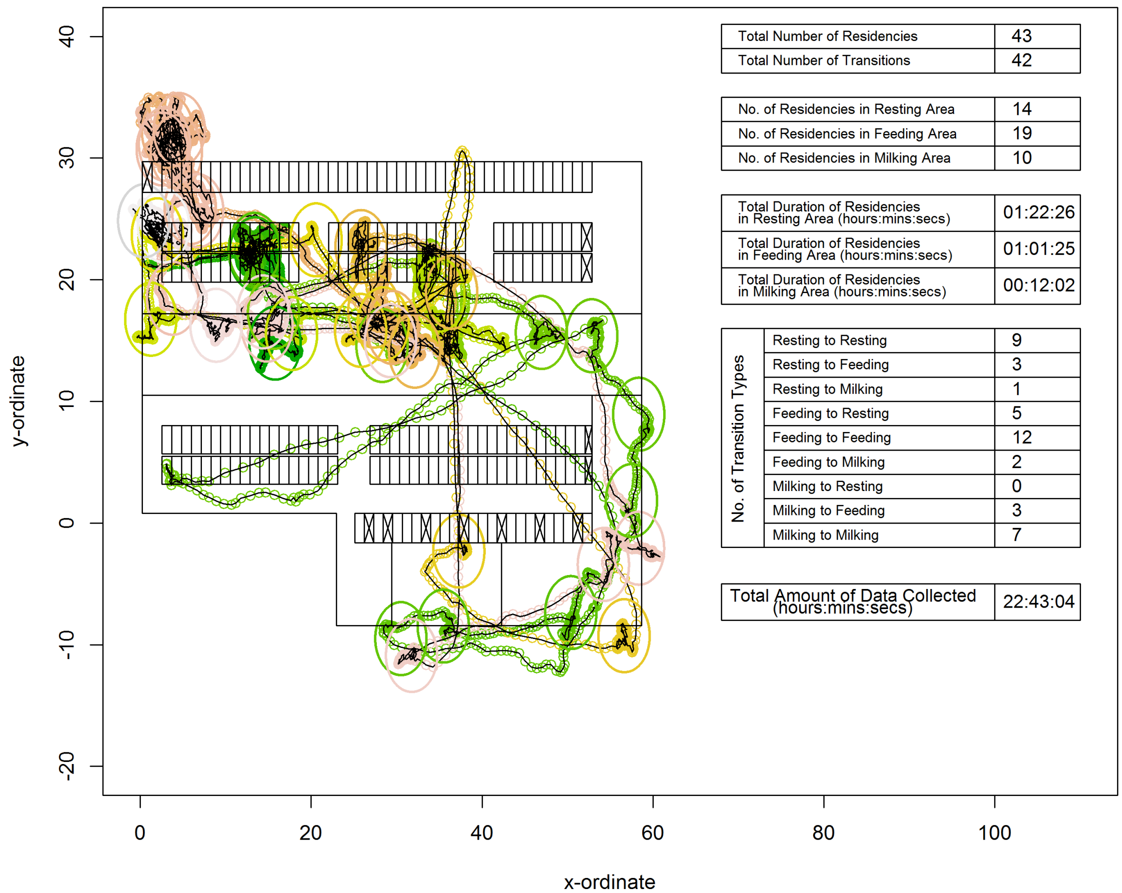
Cow 1340 Day 5 - Out-of-Control Window Size of 11 time points



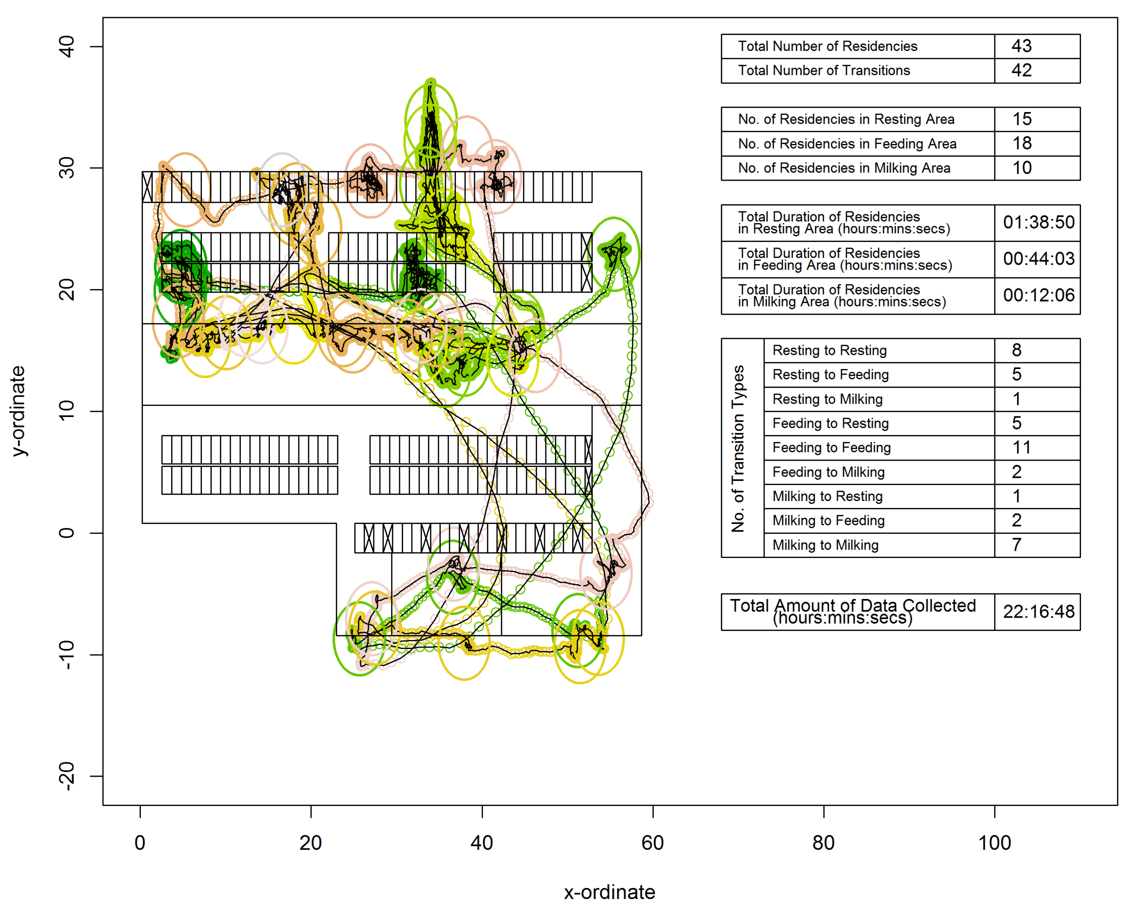
Cow 1340 Day 6 - Out-of-Control Window Size of 11 time points



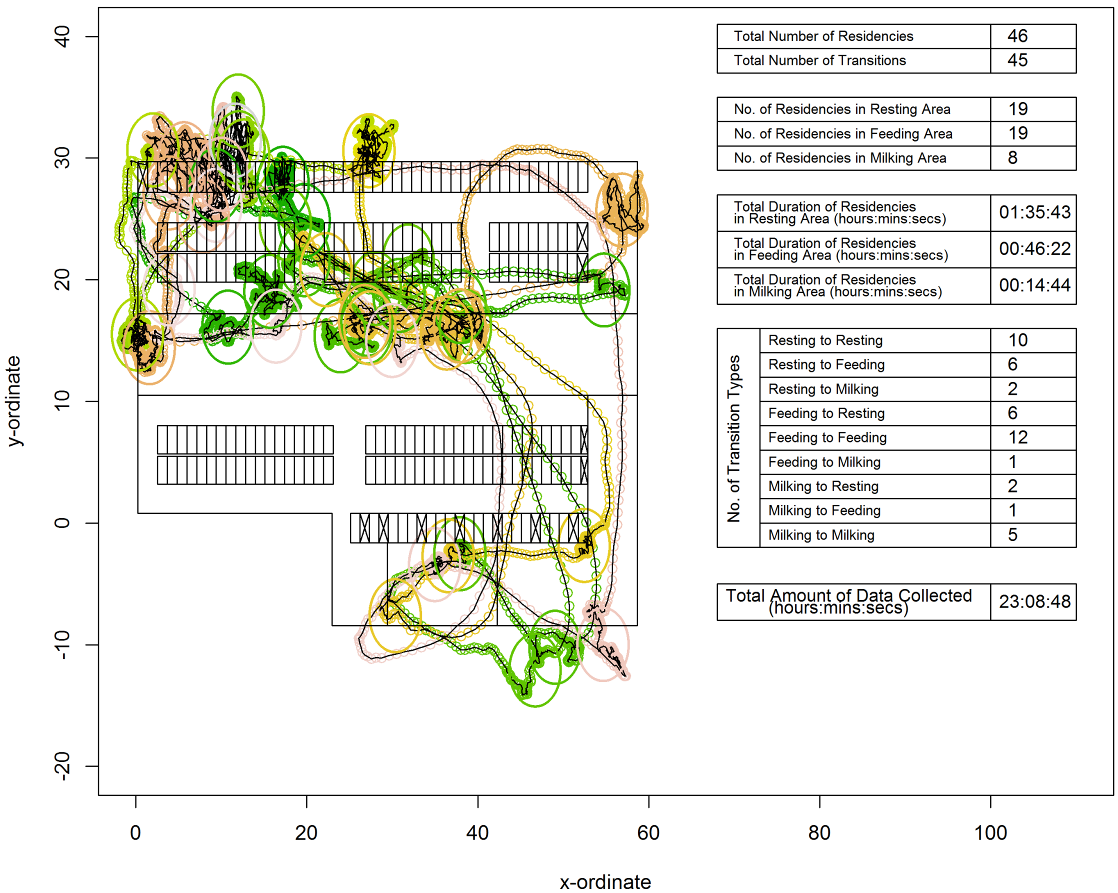
Cow 2959 Day 2 - Out-of-Control Window Size of 11 time points



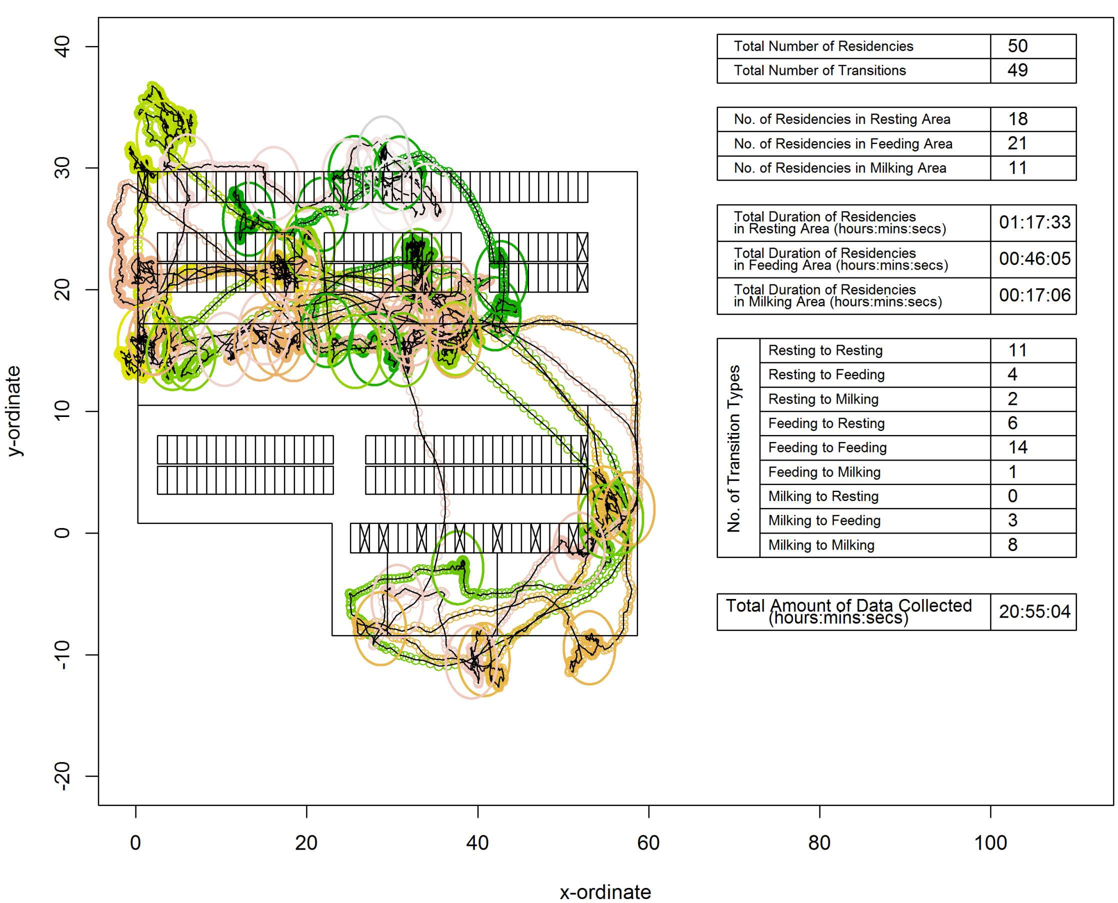
Cow 2959 Day 3 - Out-of-Control Window Size of 11 time points



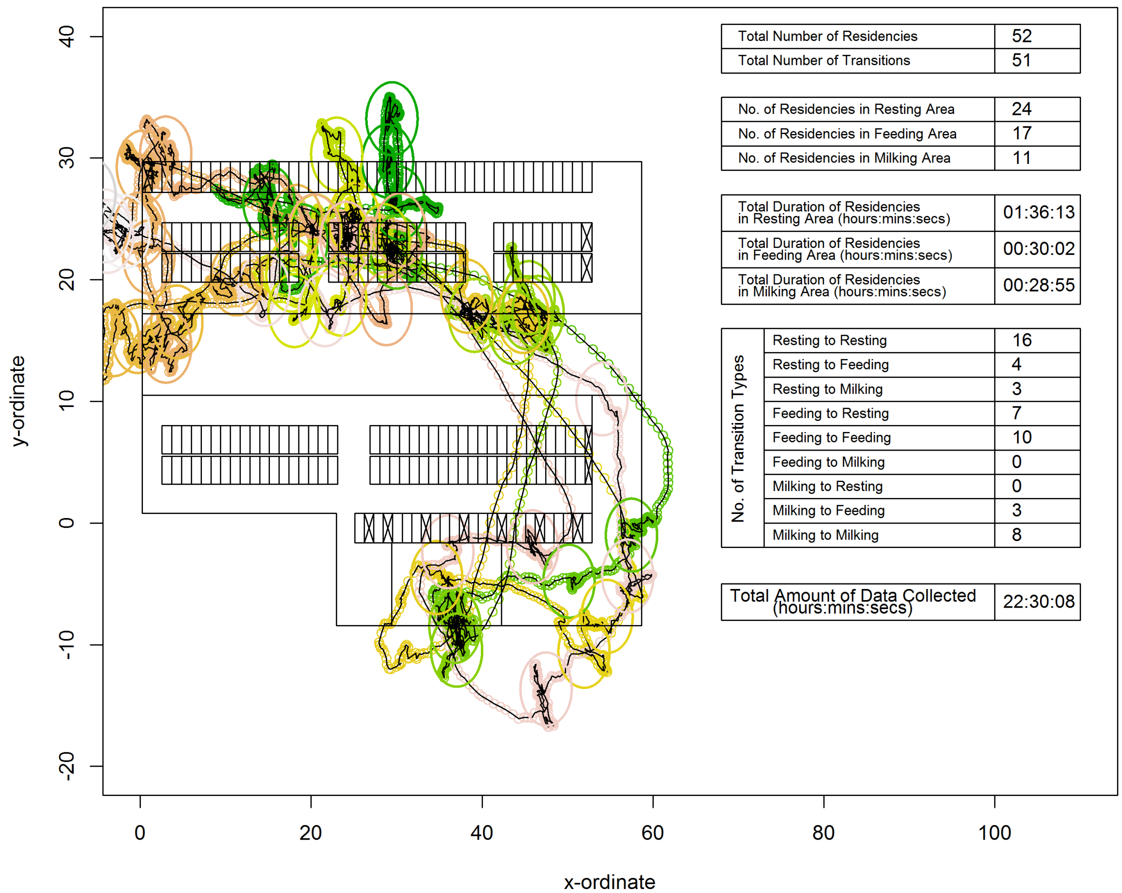
Cow 2959 Day 4 - Out-of-Control Window Size of 11 time points



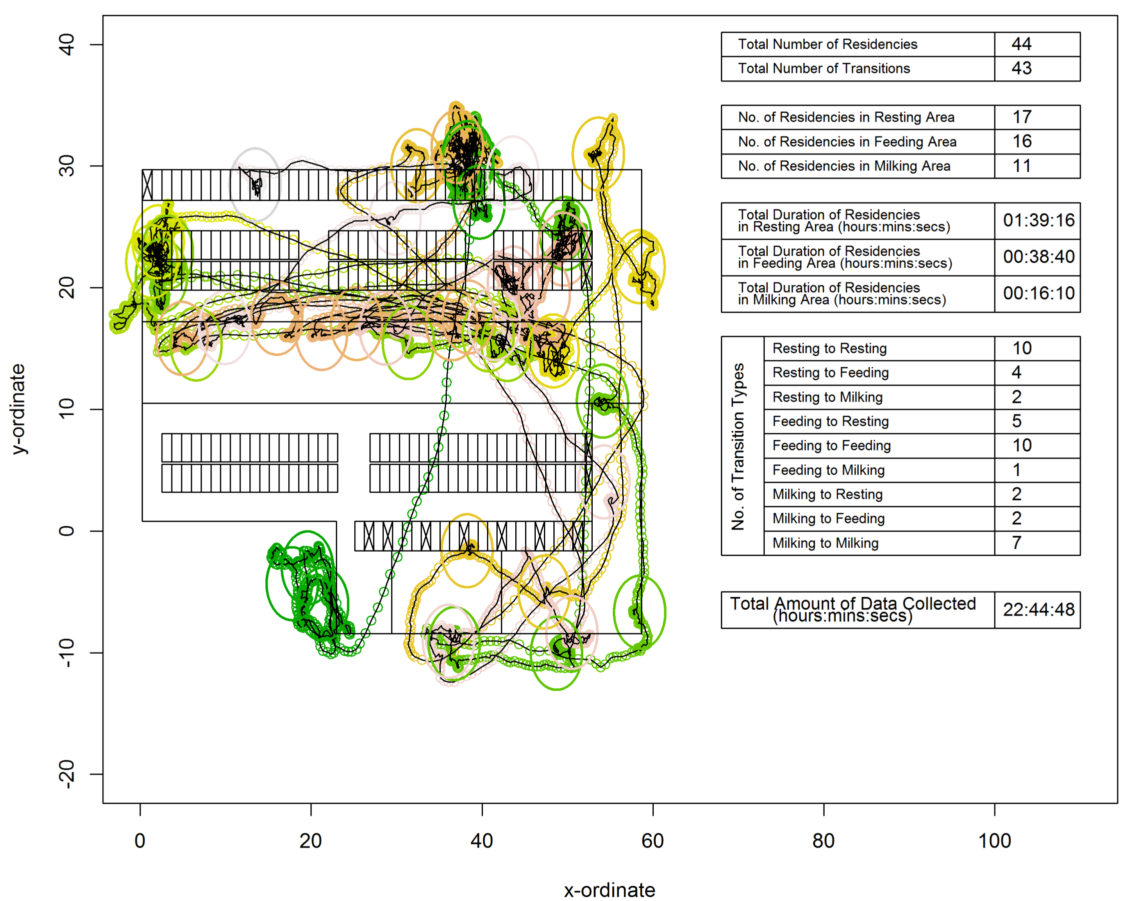
Cow 2959 Day 5 - Out-of-Control Window Size of 11 time points



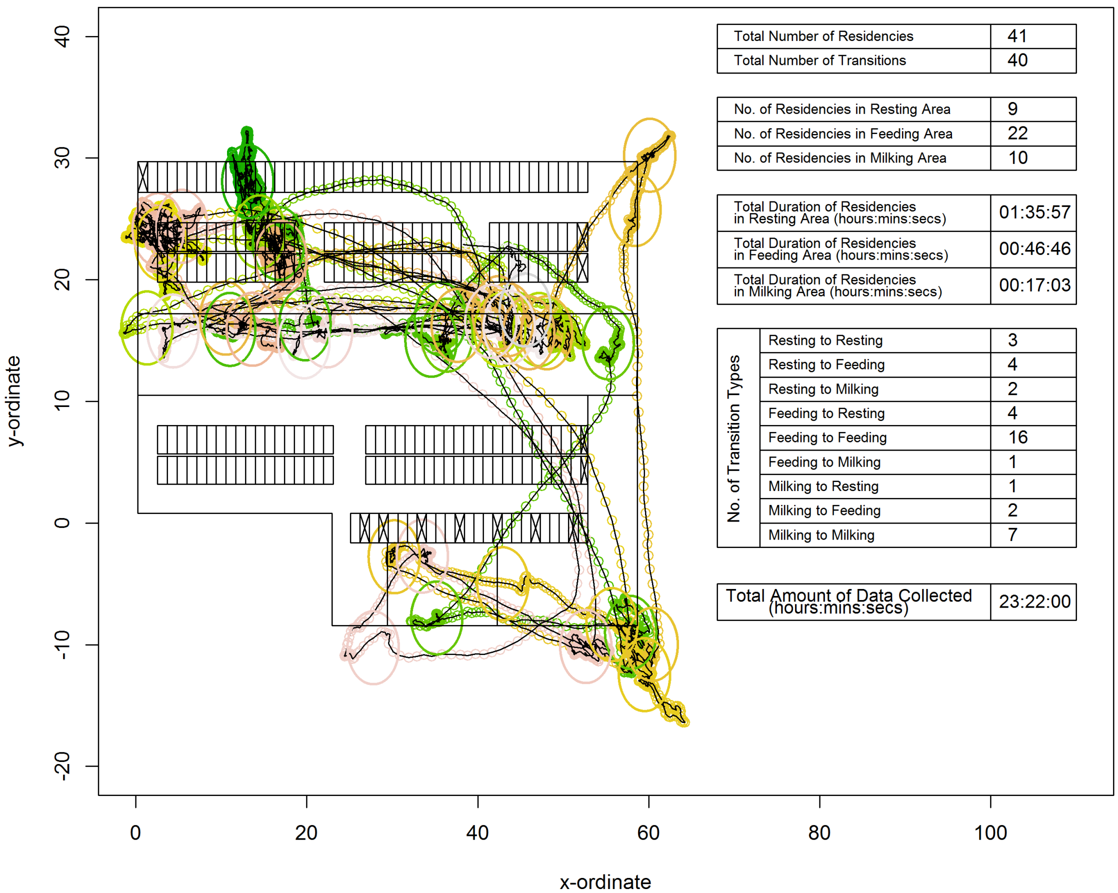
Cow 2959 Day 6 - Out-of-Control Window Size of 11 time points



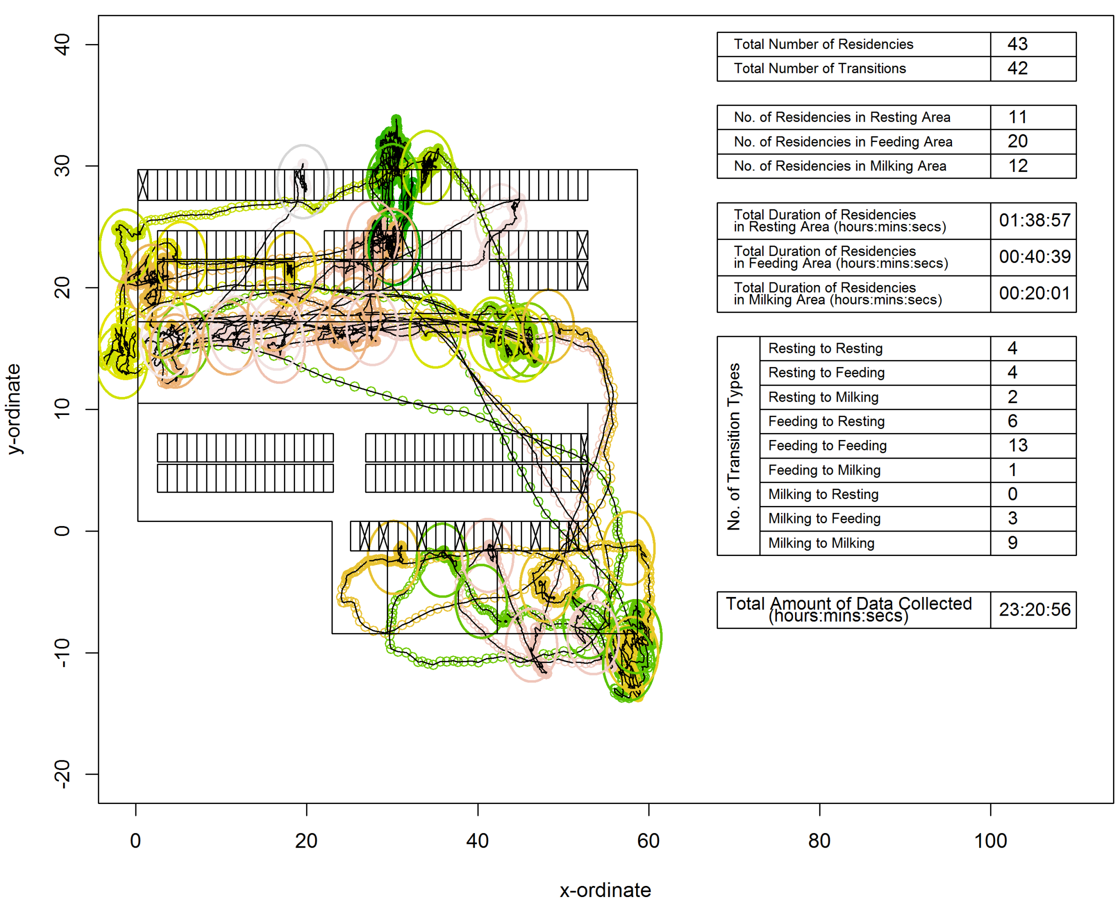
Cow 2512 Day 2 - Out-of-Control Window Size of 11 time points



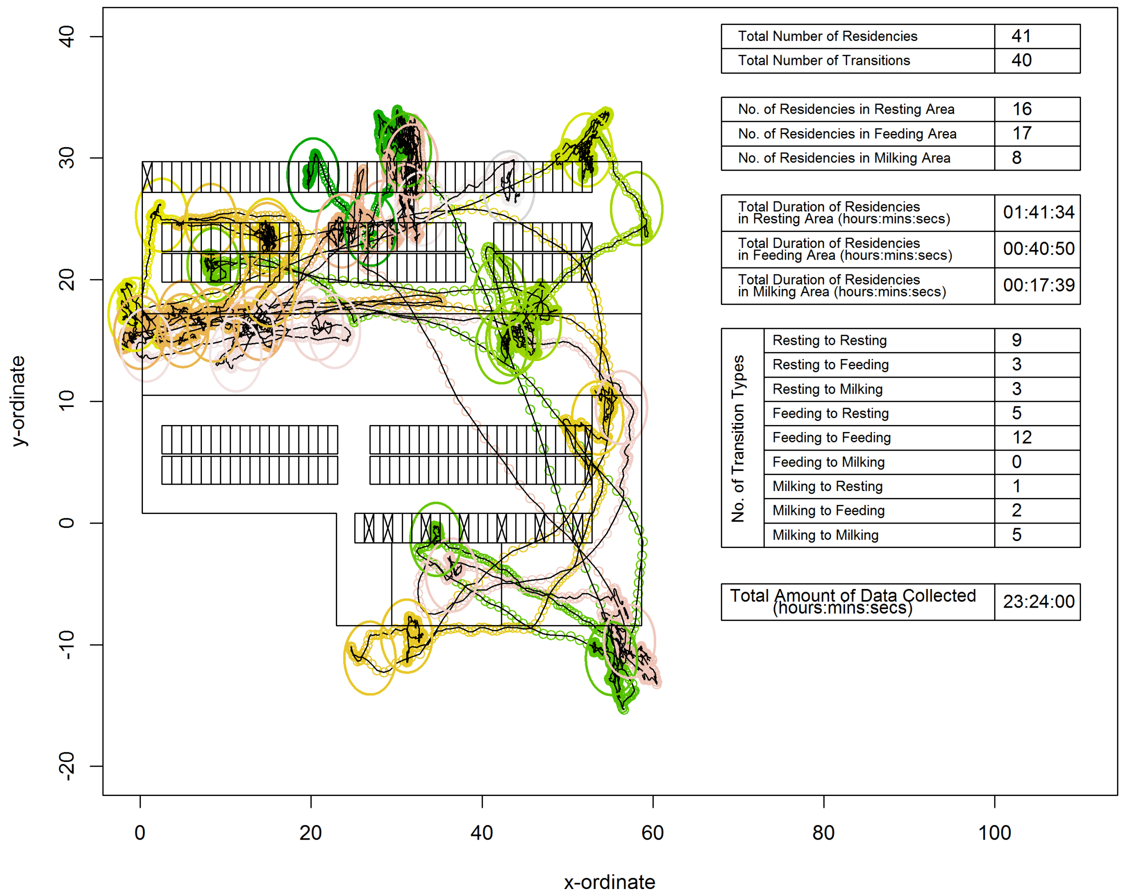
Cow 2512 Day 3 - Out-of-Control Window Size of 11 time points



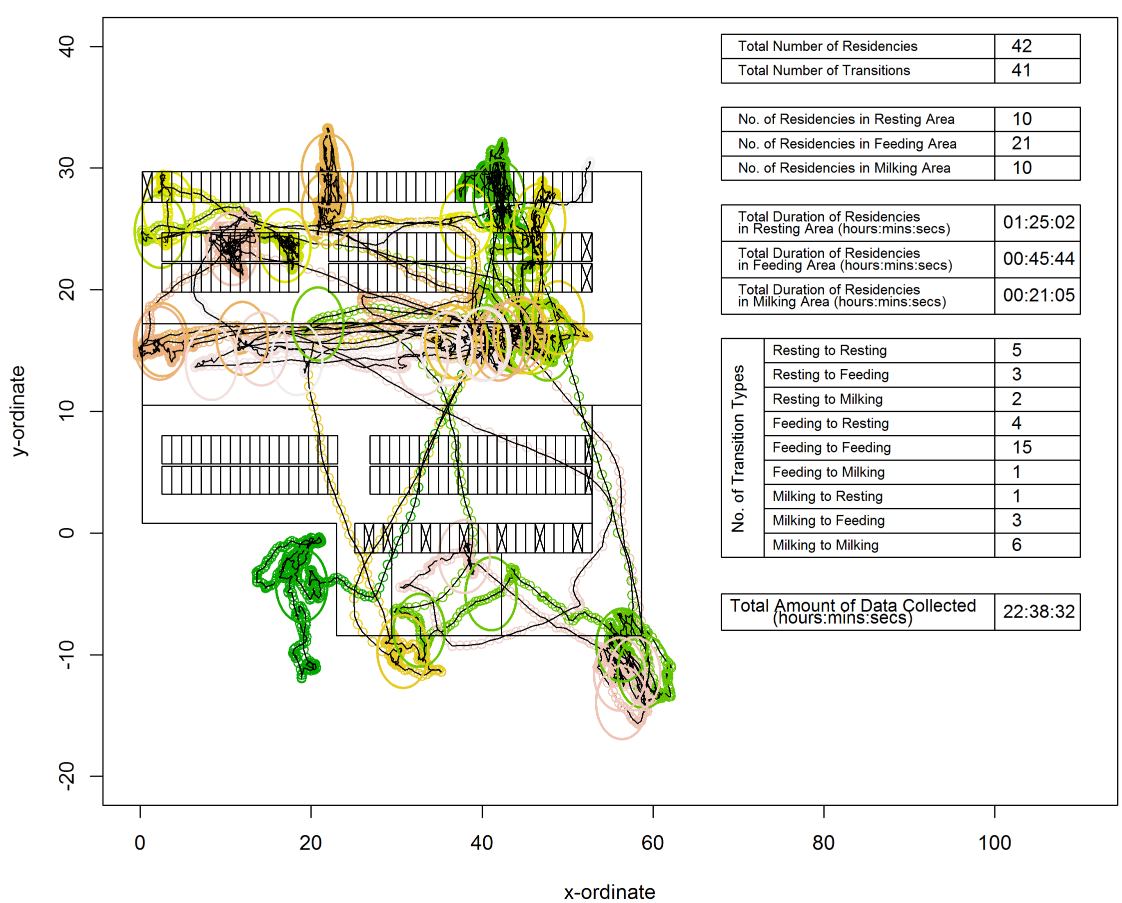
Cow 2512 Day 4 - Out-of-Control Window Size of 11 time points



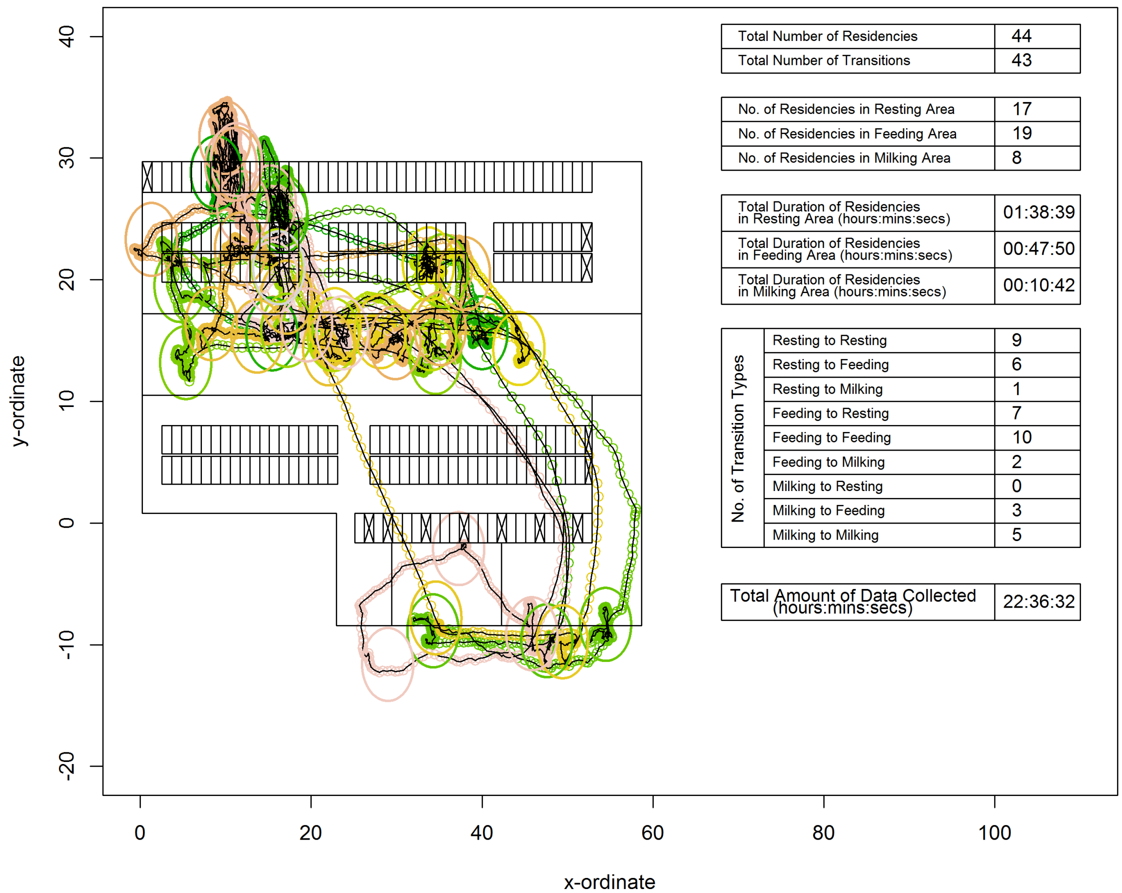
Cow 2512 Day 5 - Out-of-Control Window Size of 11 time points



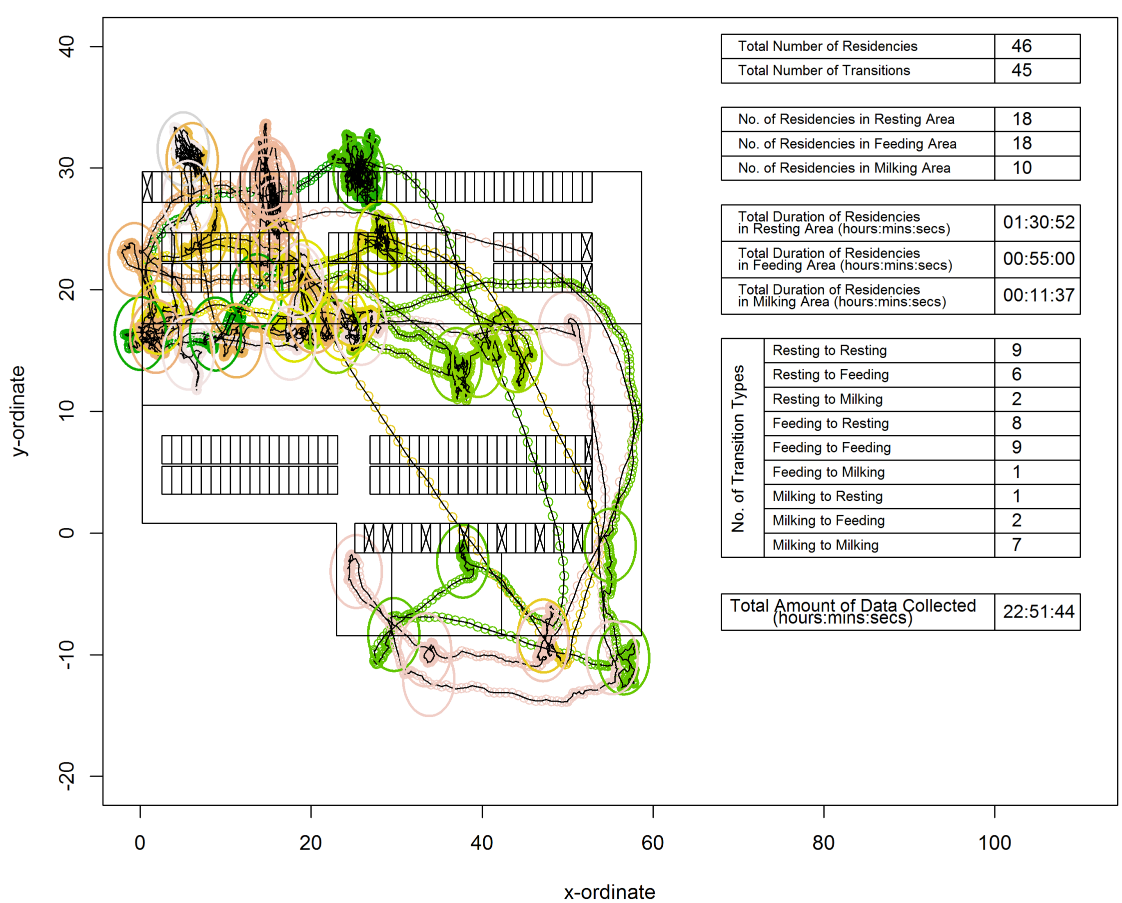
Cow 2512 Day 6 - Out-of-Control Window Size of 11 time points



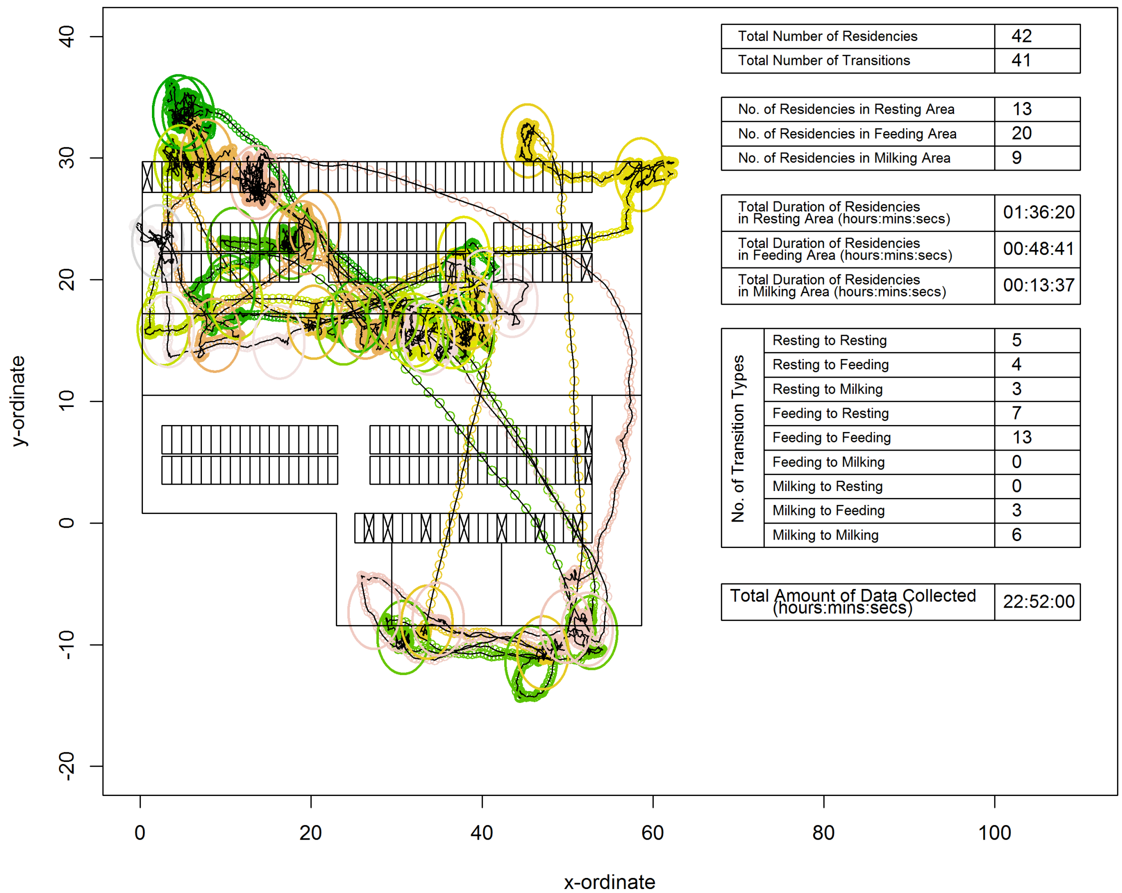
Cow 2472 Day 2 - Out-of-Control Window Size of 11 time points



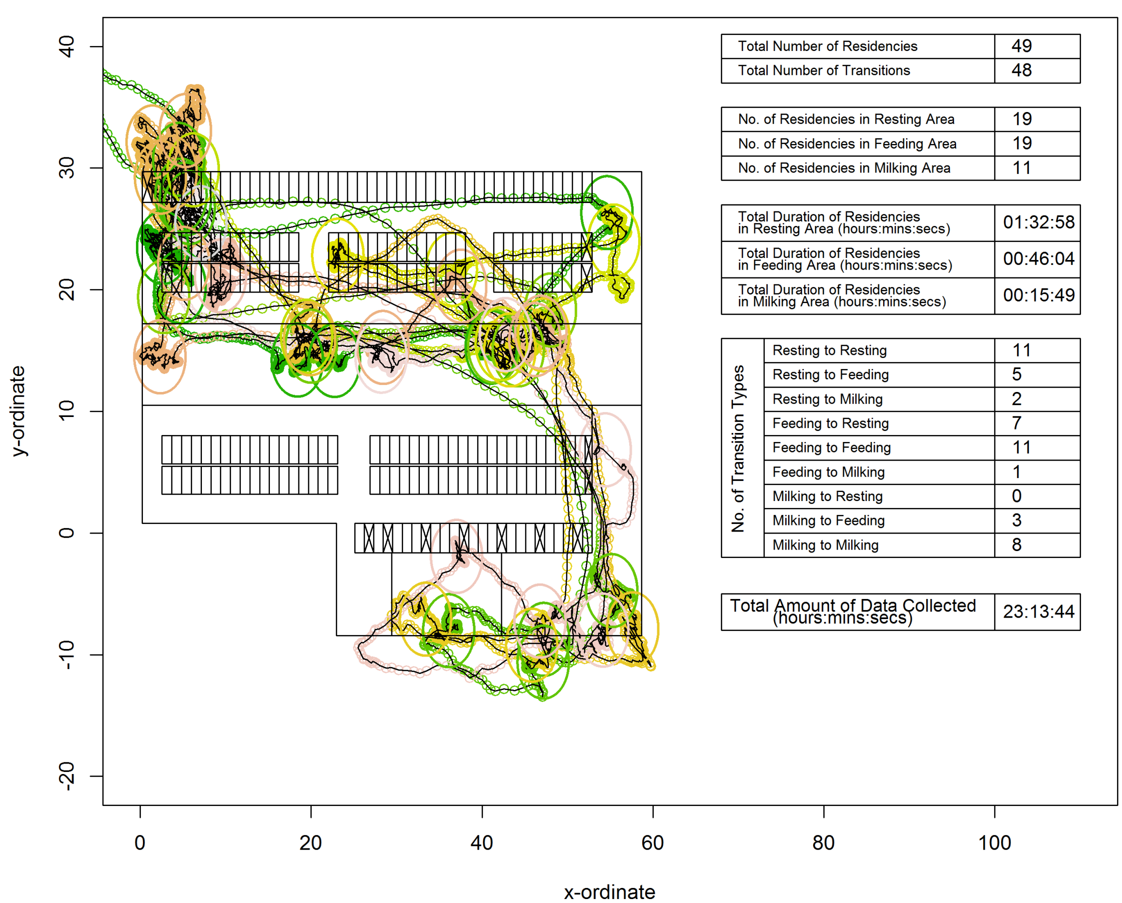
Cow 2472 Day 3 - Out-of-Control Window Size of 11 time points



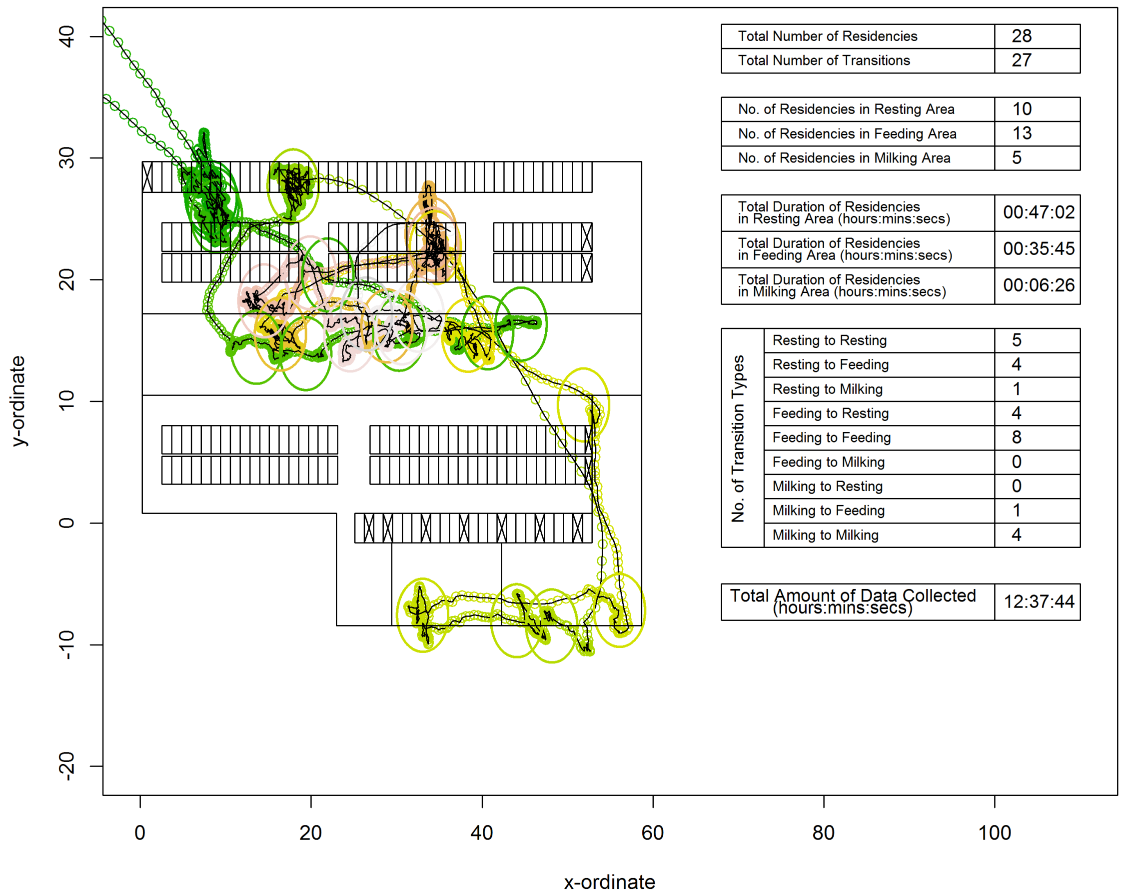
Cow 2472 Day 4 - Out-of-Control Window Size of 11 time points



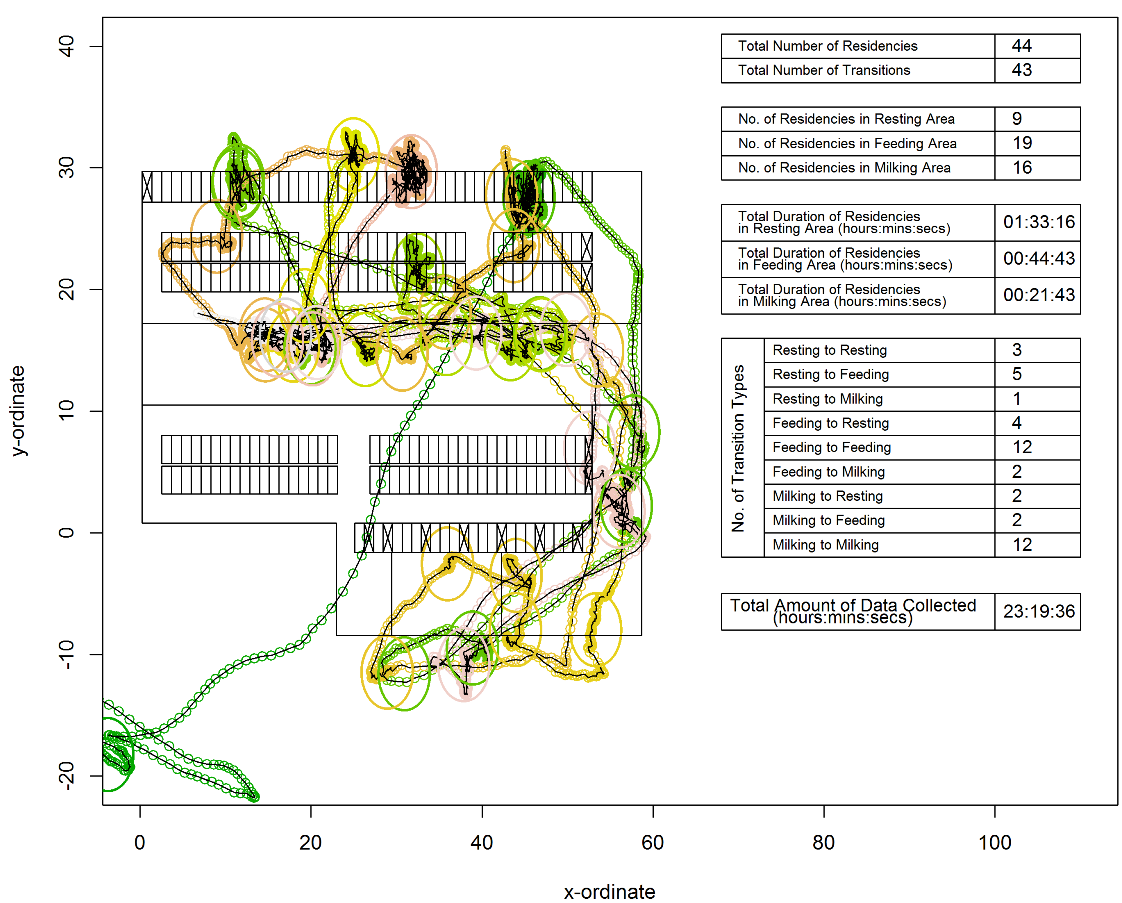
Cow 2472 Day 5 - Out-of-Control Window Size of 11 time points



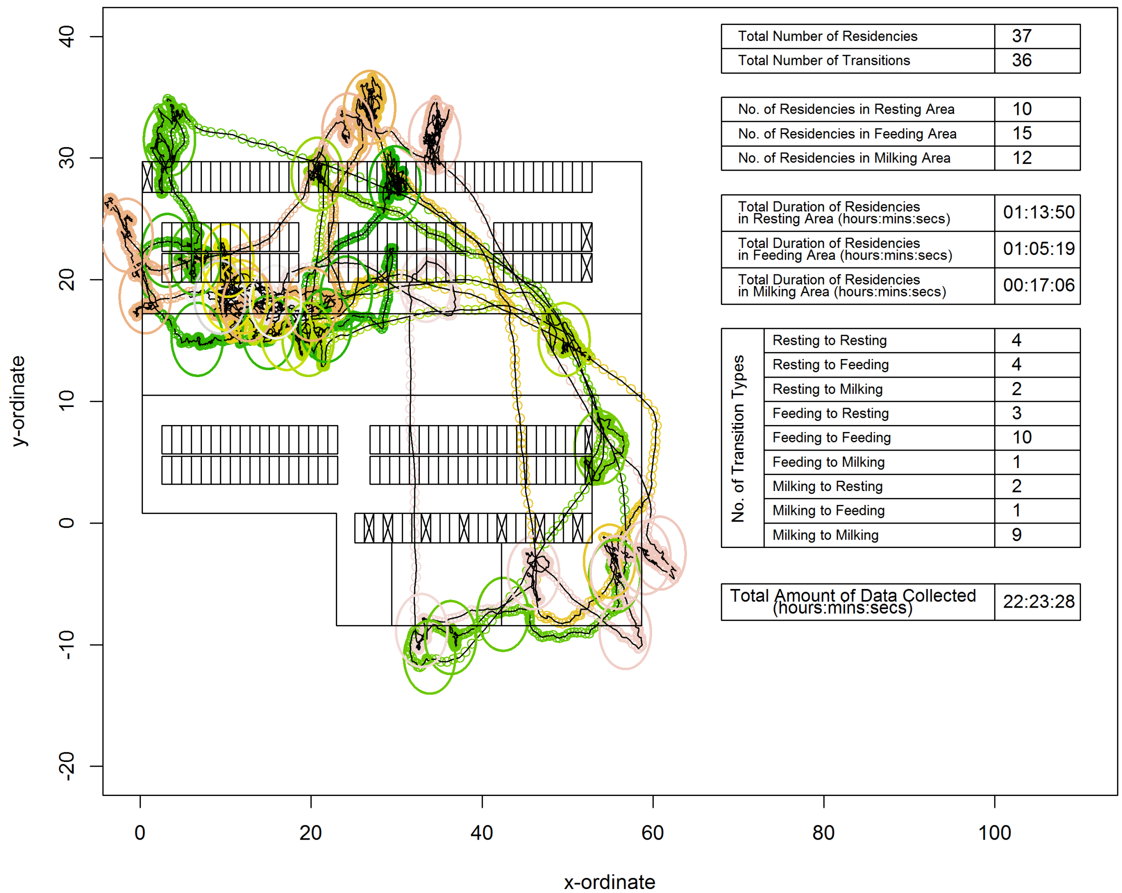
Cow 2472 Day 6 - Out-of-Control Window Size of 11 time points



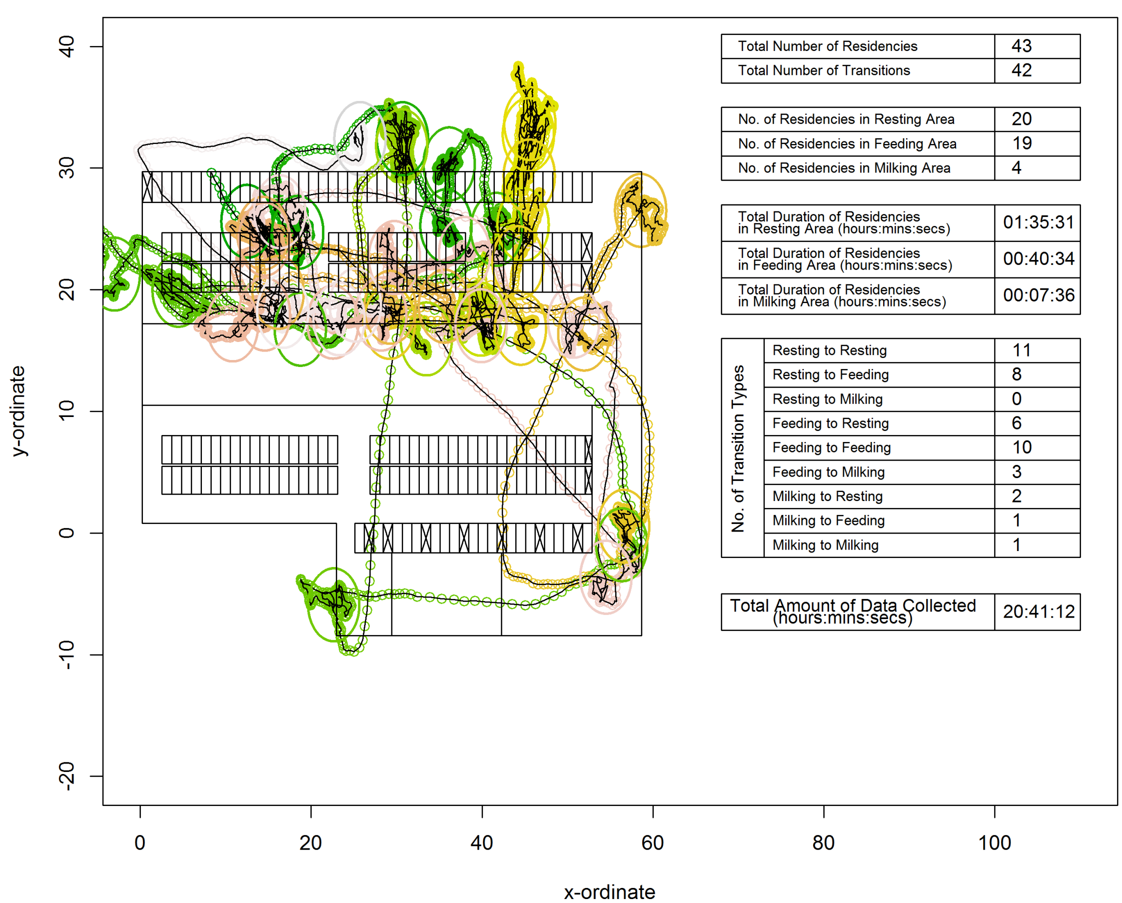
Cow 2179 Day 2 - Out-of-Control Window Size of 11 time points



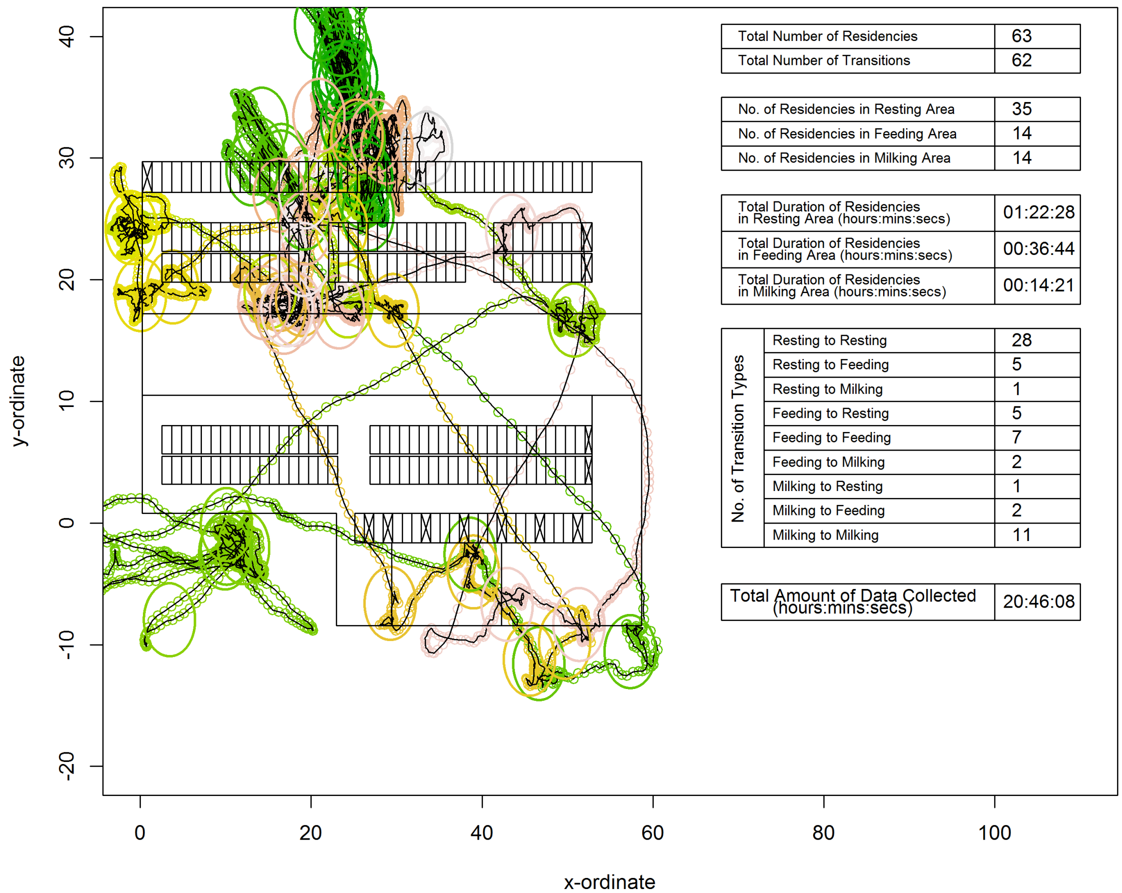
Cow 2179 Day 3 - Out-of-Control Window Size of 11 time points



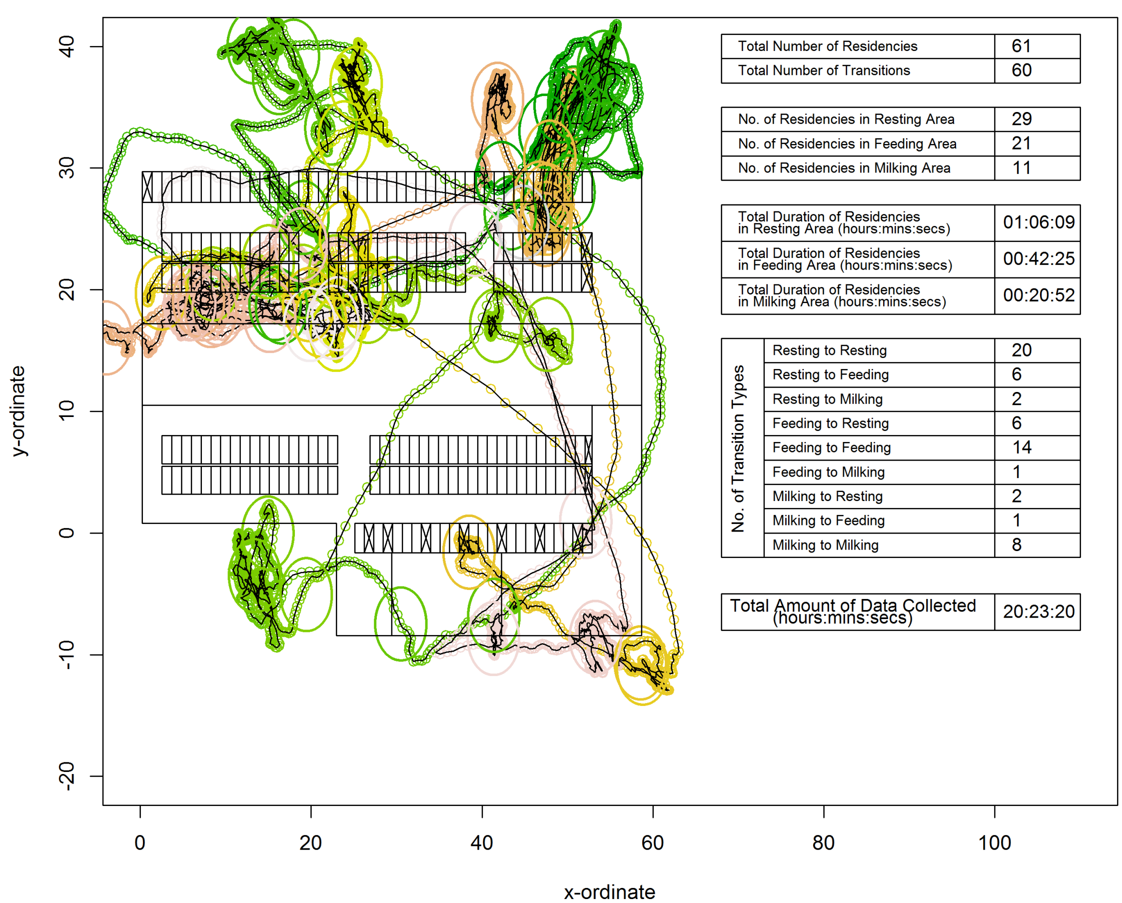
Cow 2179 Day 4 - Out-of-Control Window Size of 11 time points



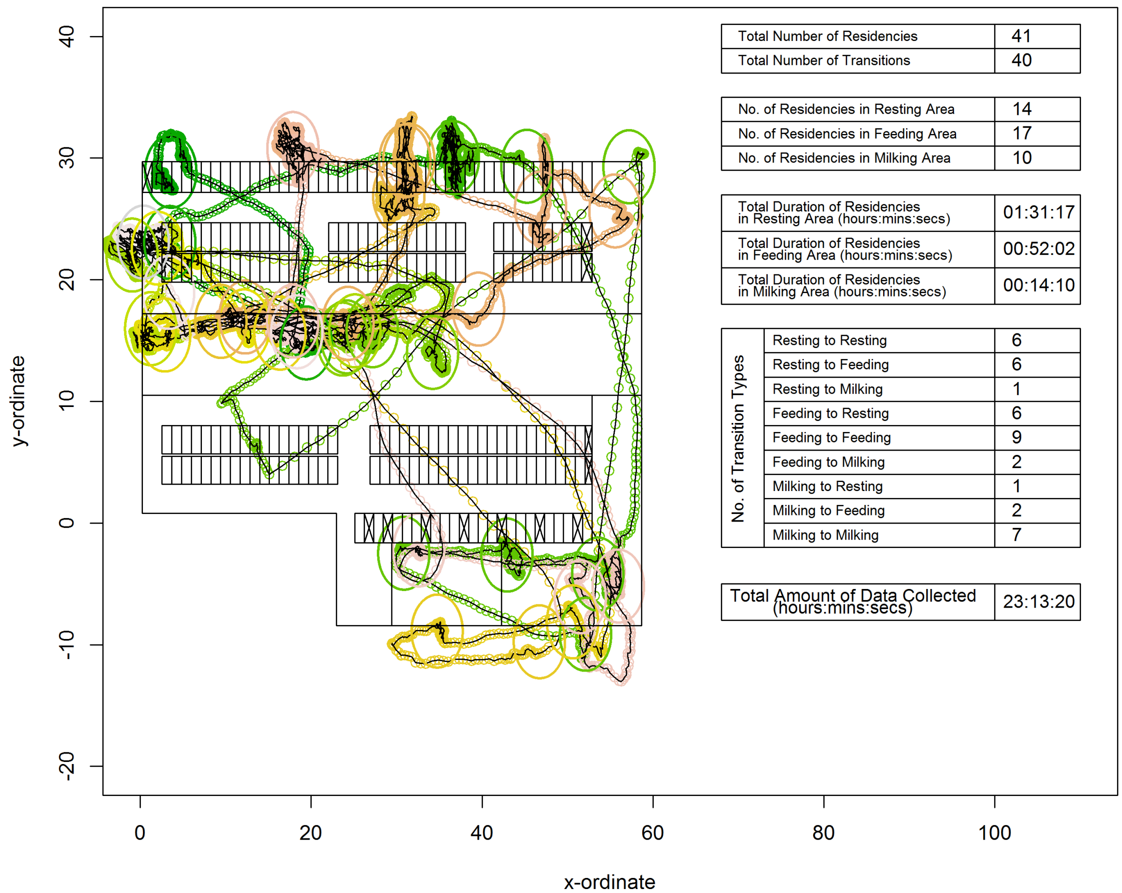
Cow 2179 Day 5 - Out-of-Control Window Size of 11 time points



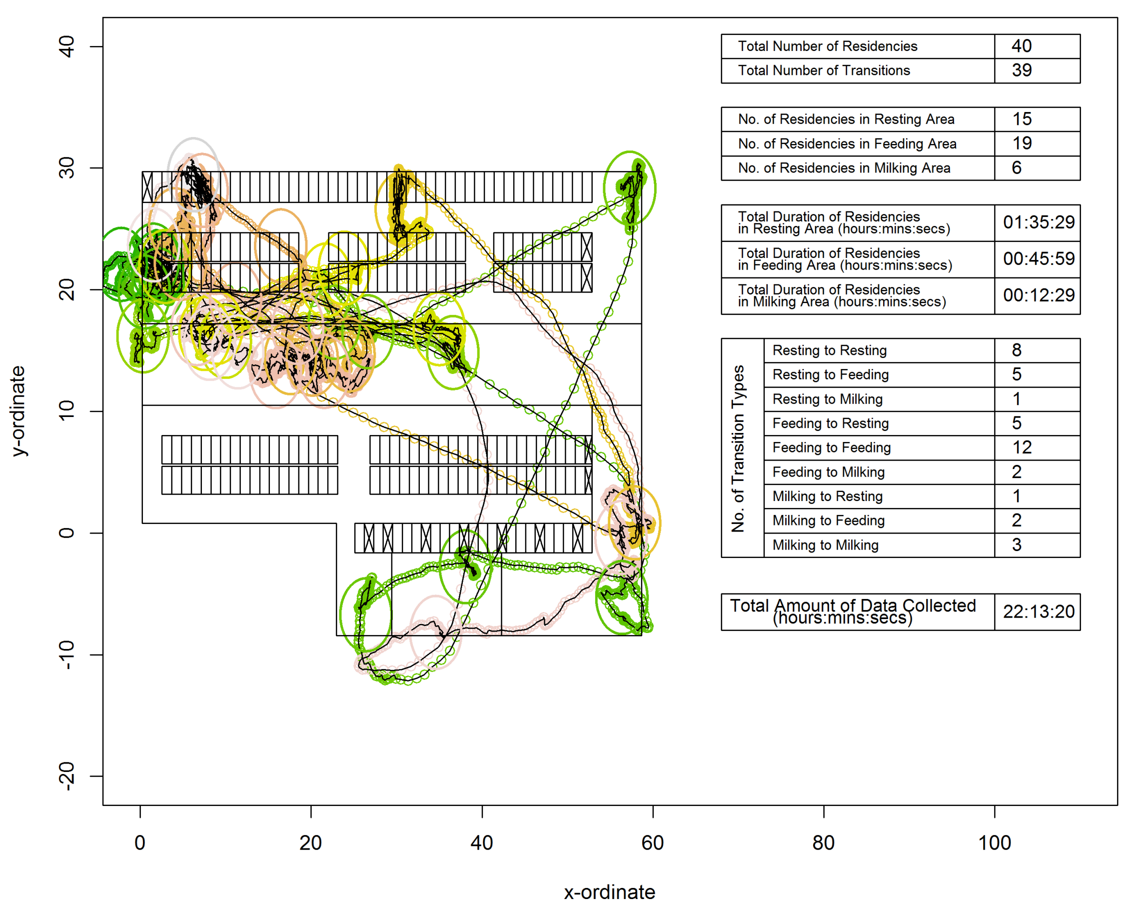
Cow 2179 Day 6 - Out-of-Control Window Size of 11 time points



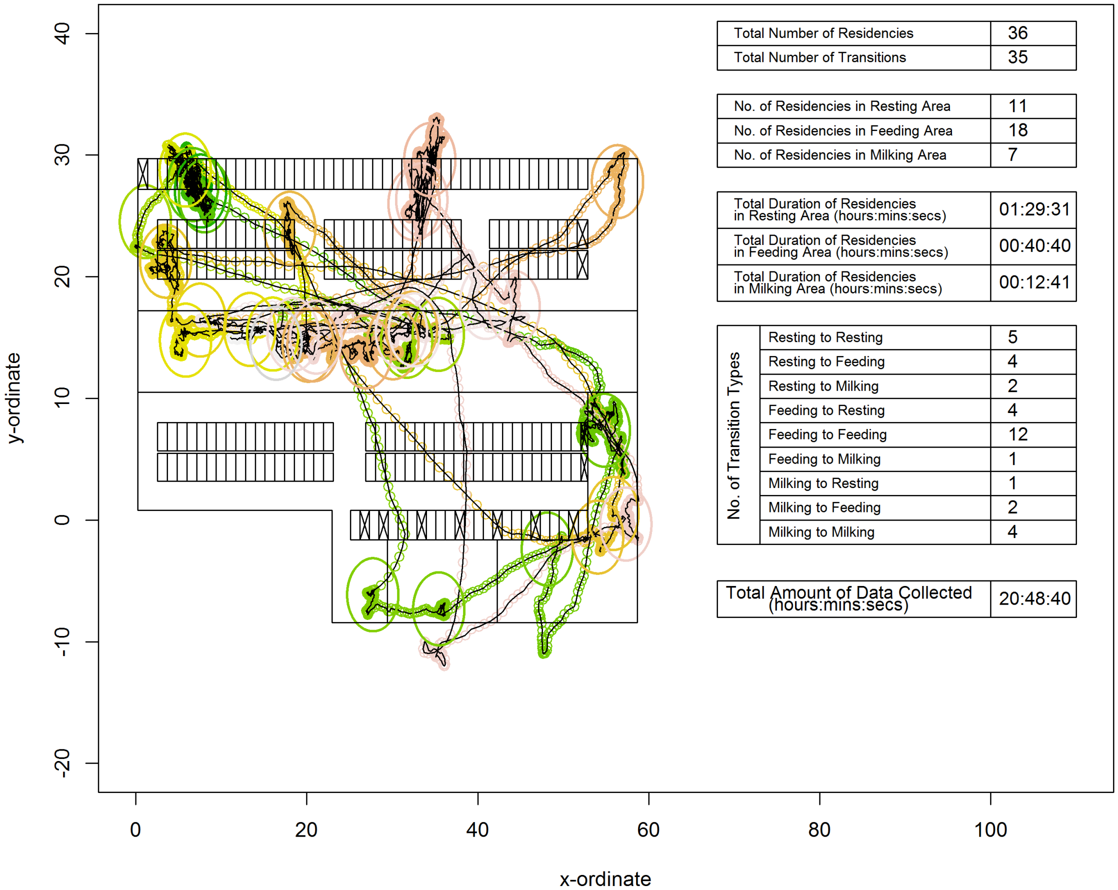
Cow 2172 Day 2 - Out-of-Control Window Size of 11 time points



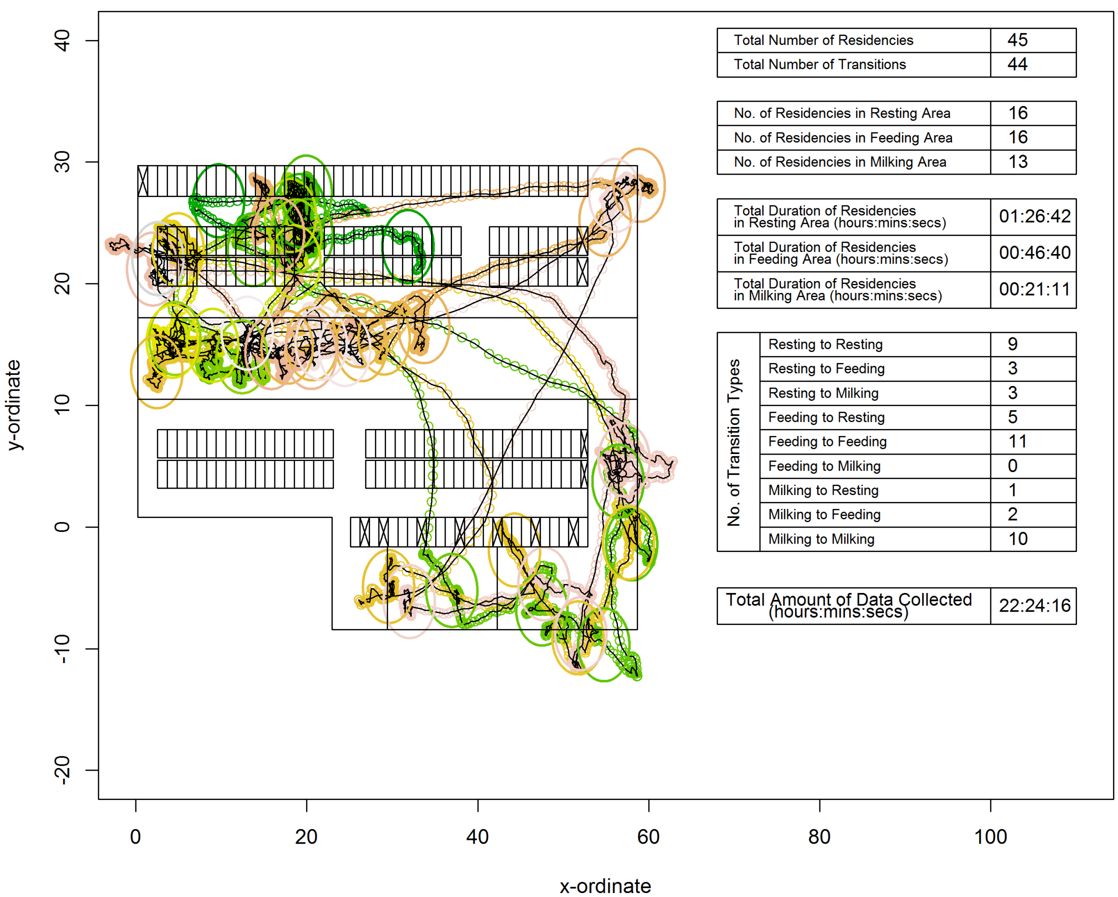
Cow 2172 Day 3 - Out-of-Control Window Size of 11 time points



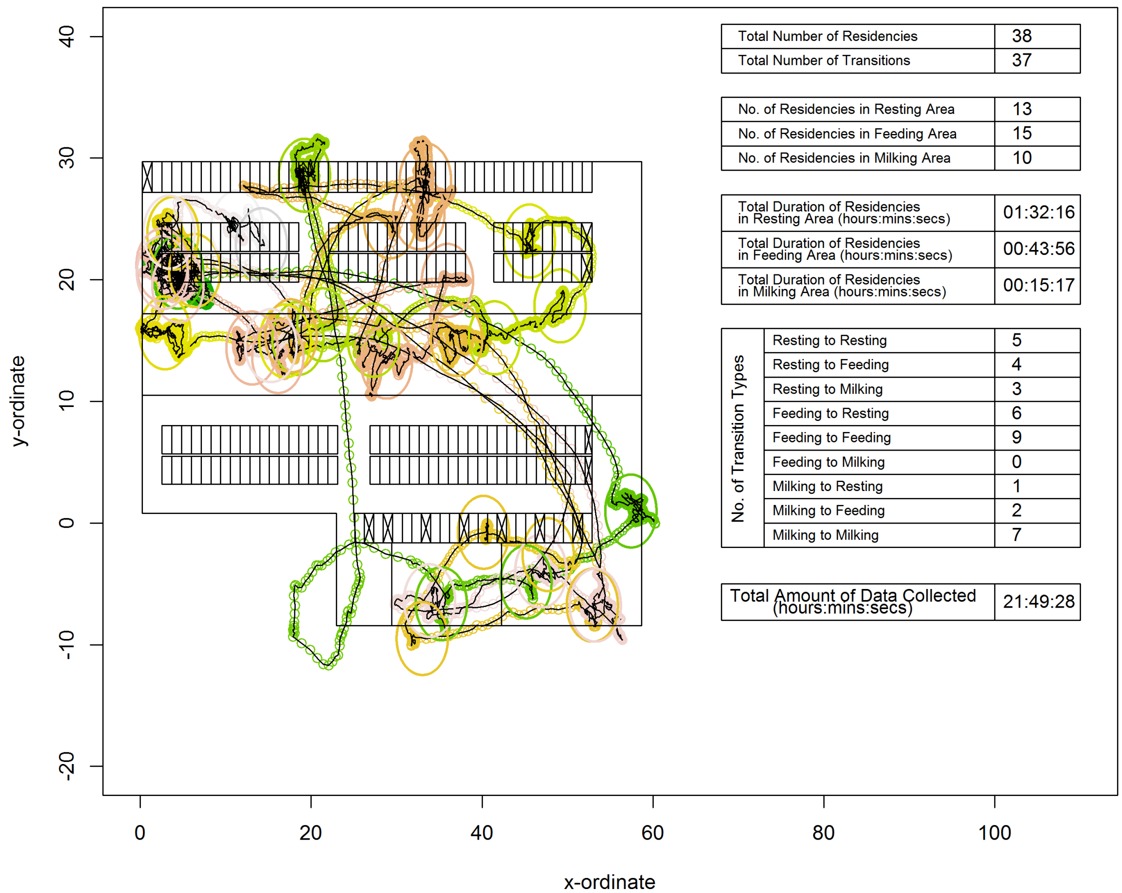
Cow 2172 Day 4 - Out-of-Control Window Size of 11 time points



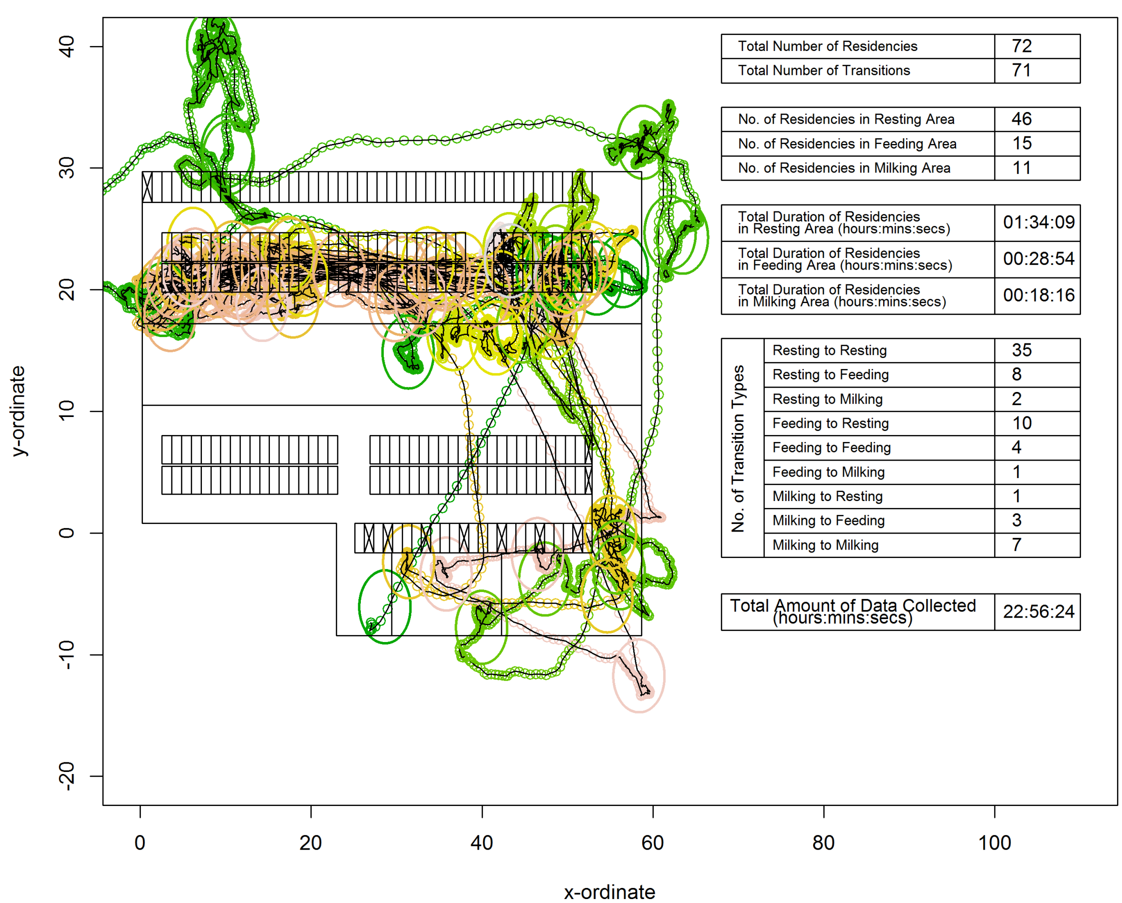
Cow 2172 Day 5 - Out-of-Control Window Size of 11 time points



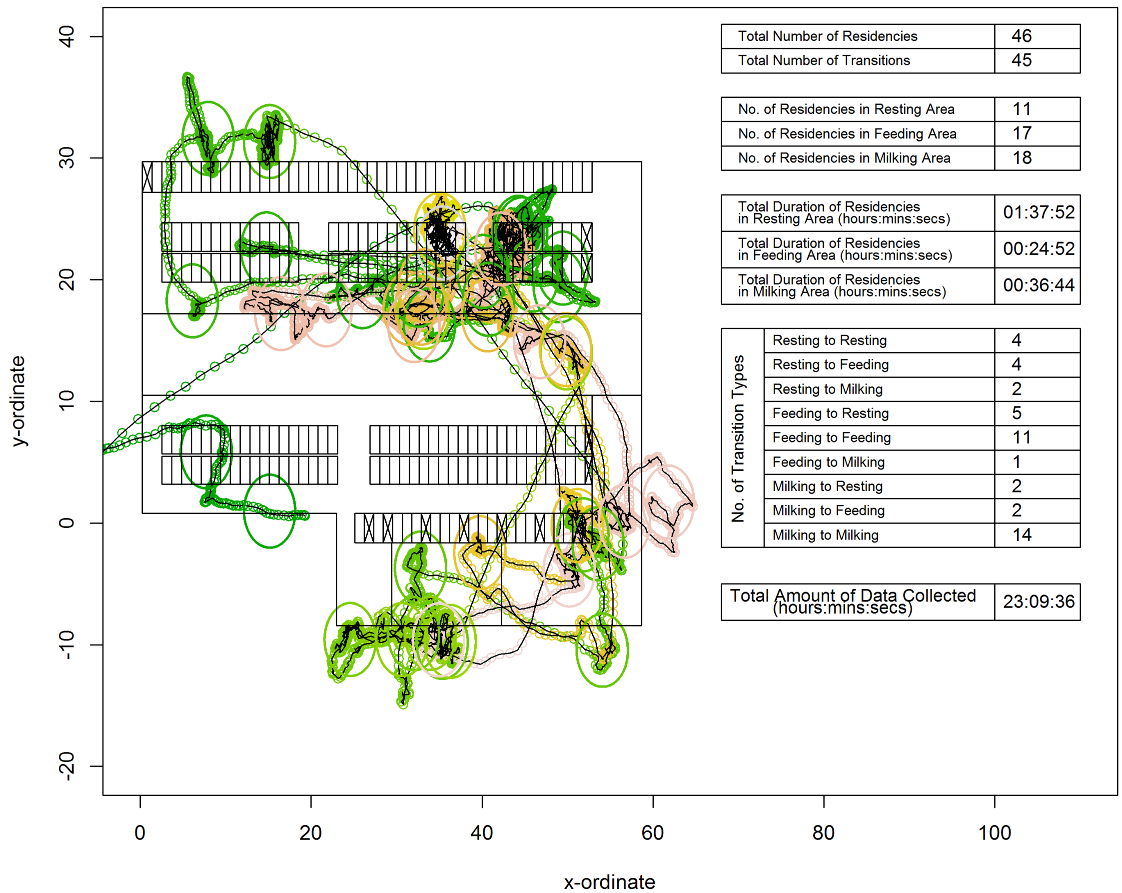
Cow 2172 Day 6 - Out-of-Control Window Size of 11 time points



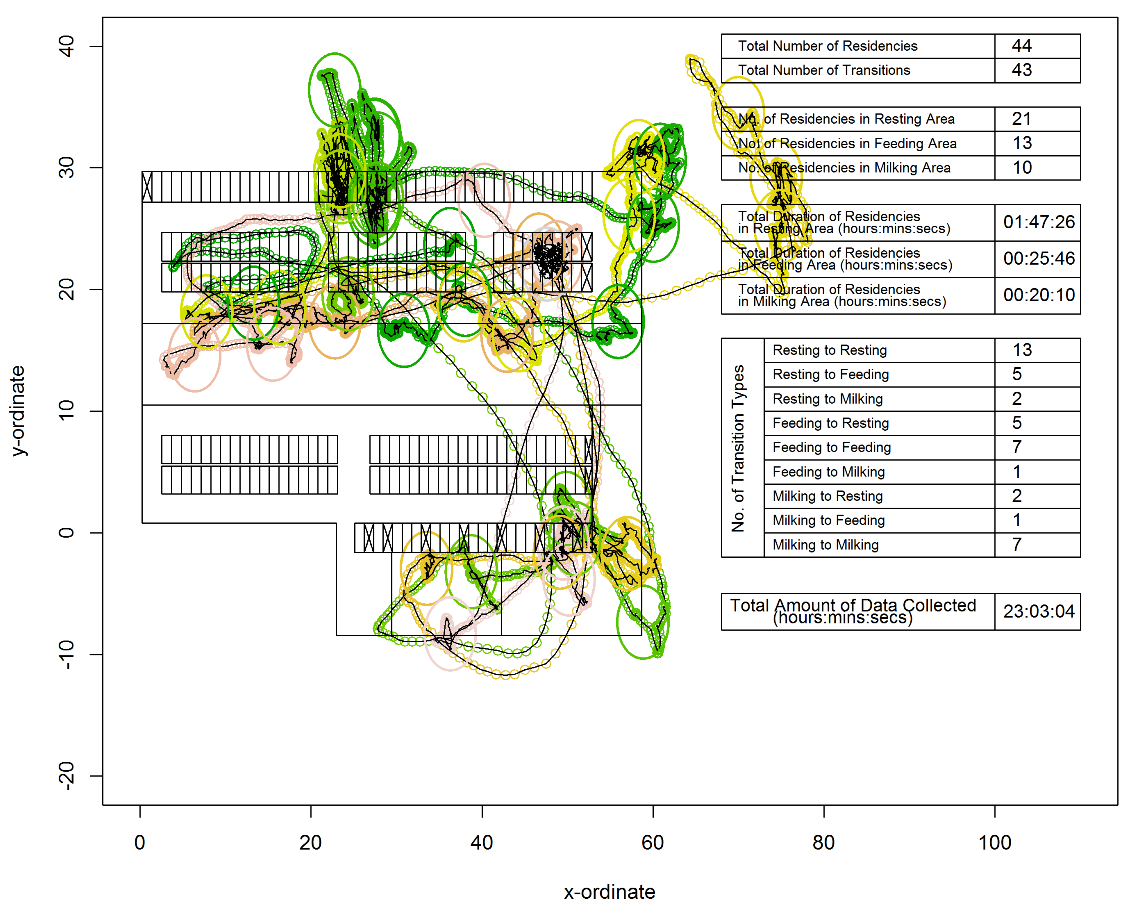
Cow 2153 Day 2 - Out-of-Control Window Size of 11 time points



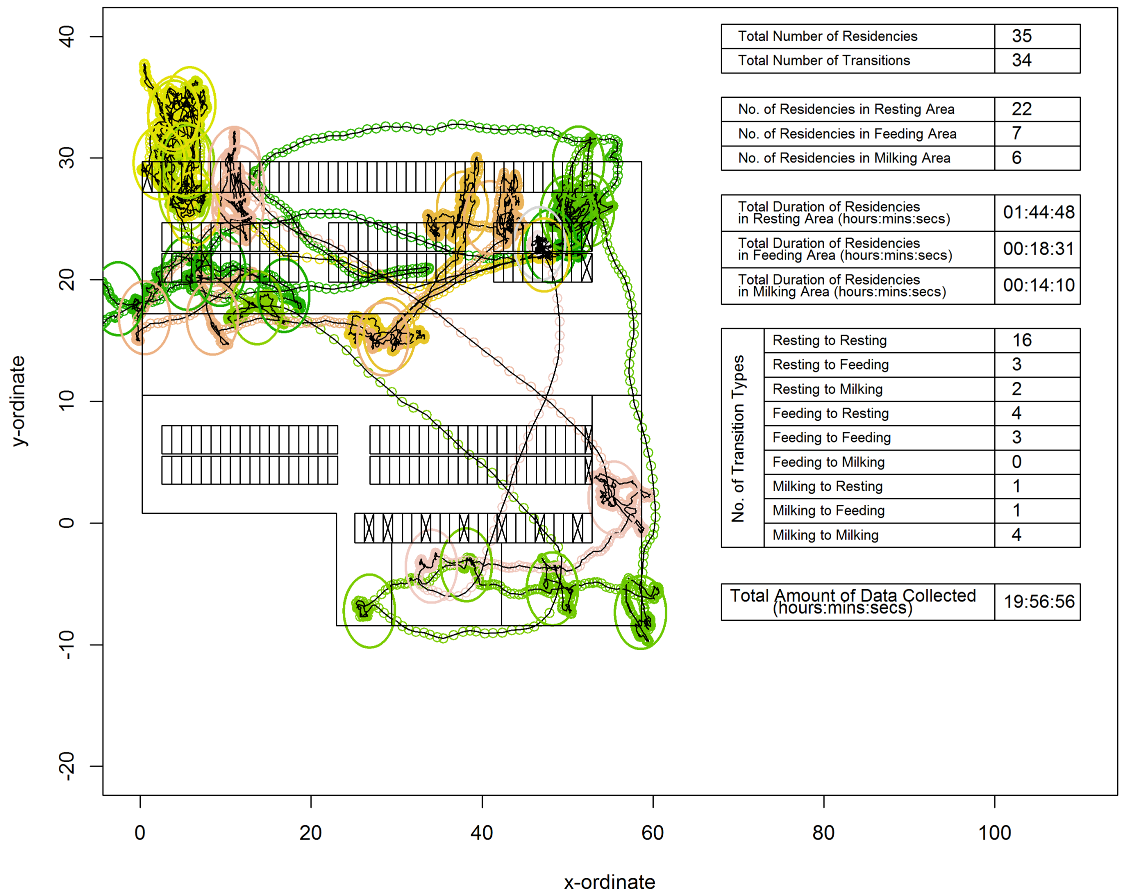
Cow 2153 Day 3 - Out-of-Control Window Size of 11 time points



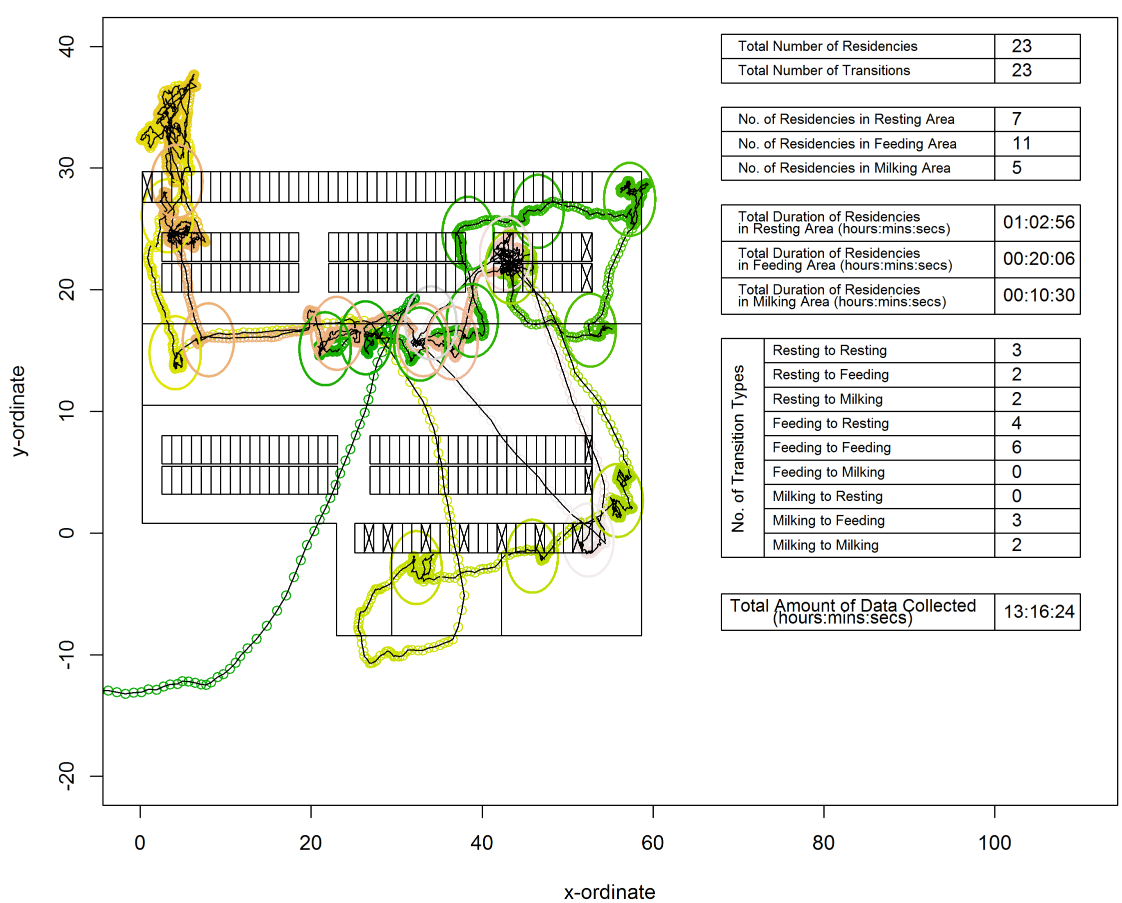
Cow 2153 Day 4 - Out-of-Control Window Size of 11 time points



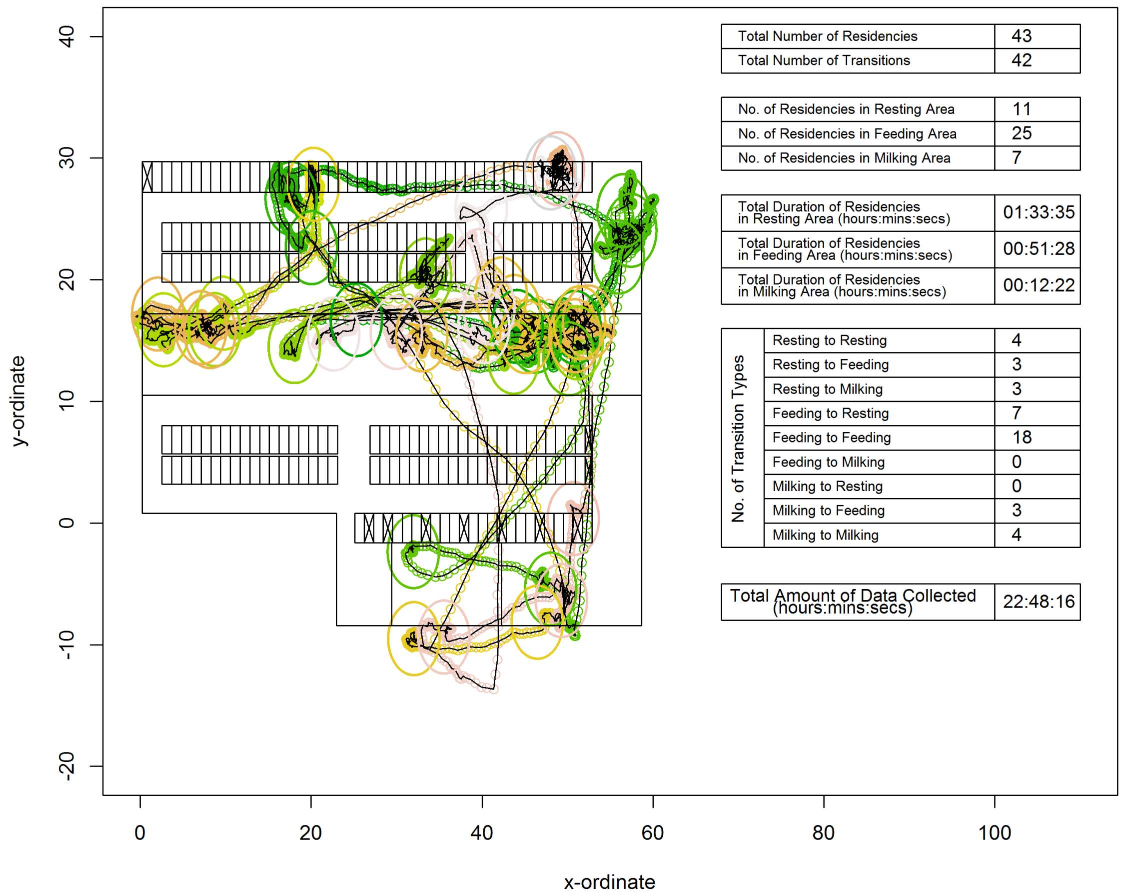
Cow 2153 Day 5 - Out-of-Control Window Size of 11 time points



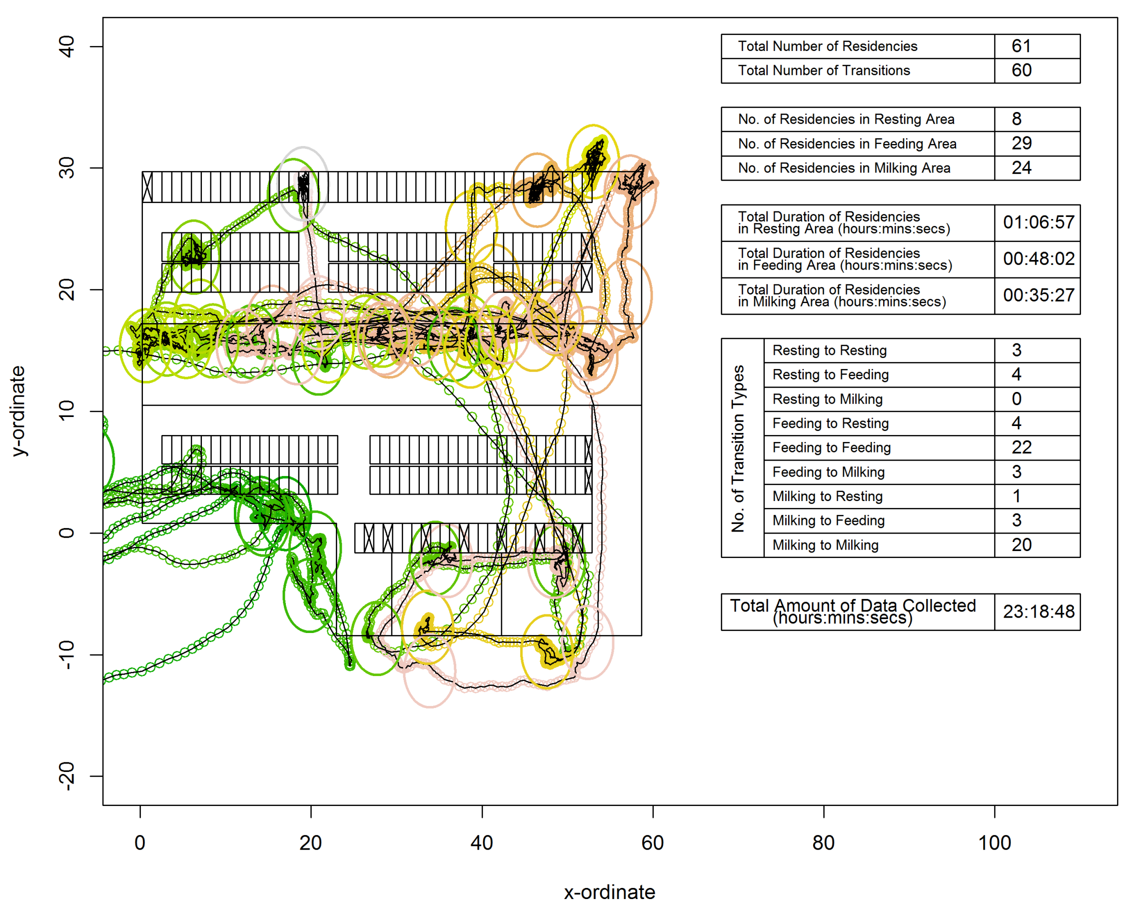
Cow 2153 Day 6 - Out-of-Control Window Size of 11 time points



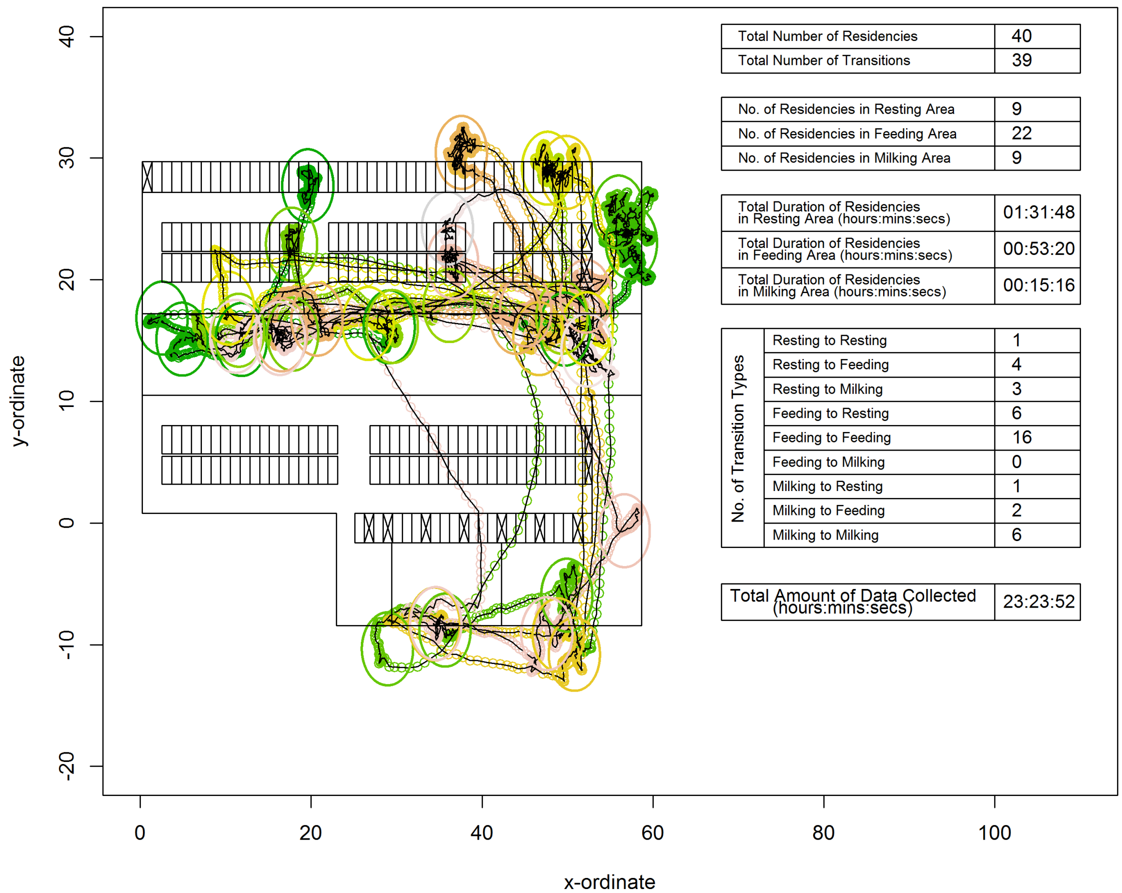
Cow 1892 Day 2 - Out-of-Control Window Size of 11 time points



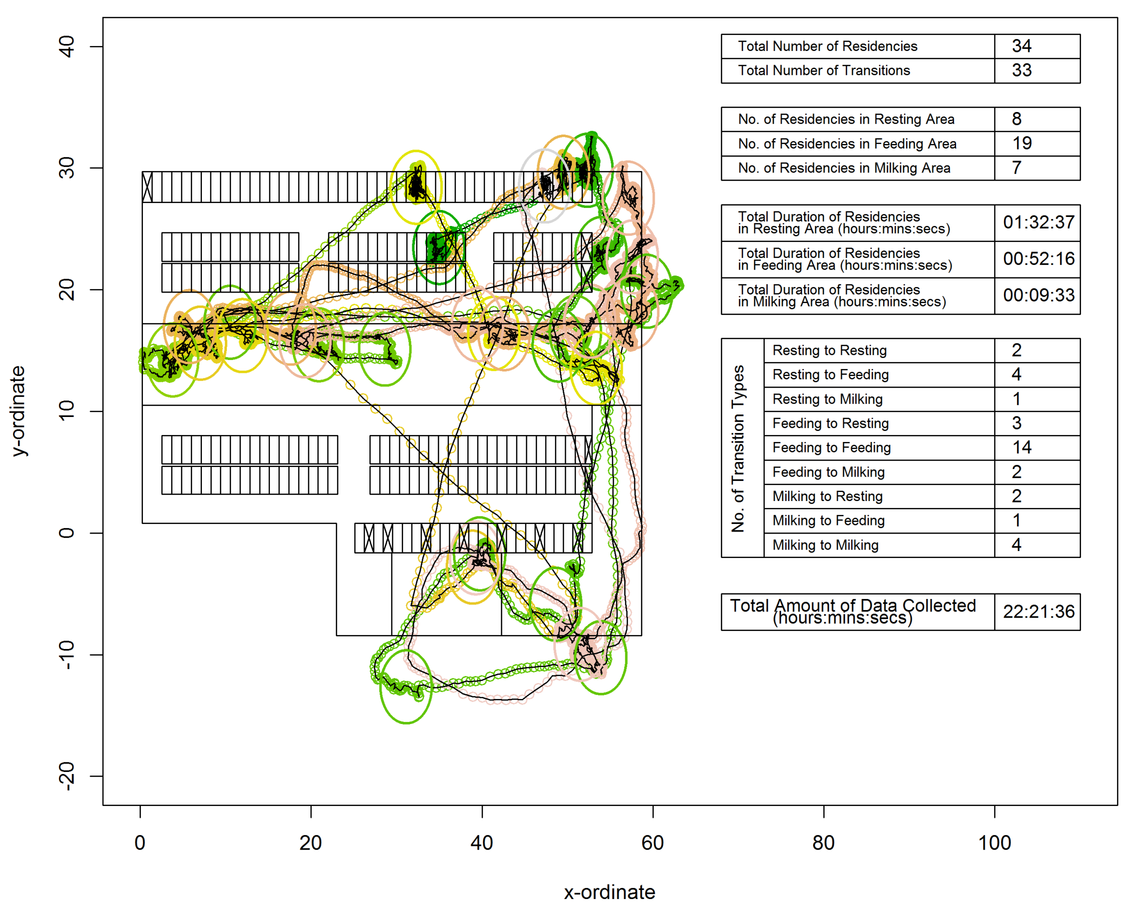
Cow 1892 Day 3 - Out-of-Control Window Size of 11 time points



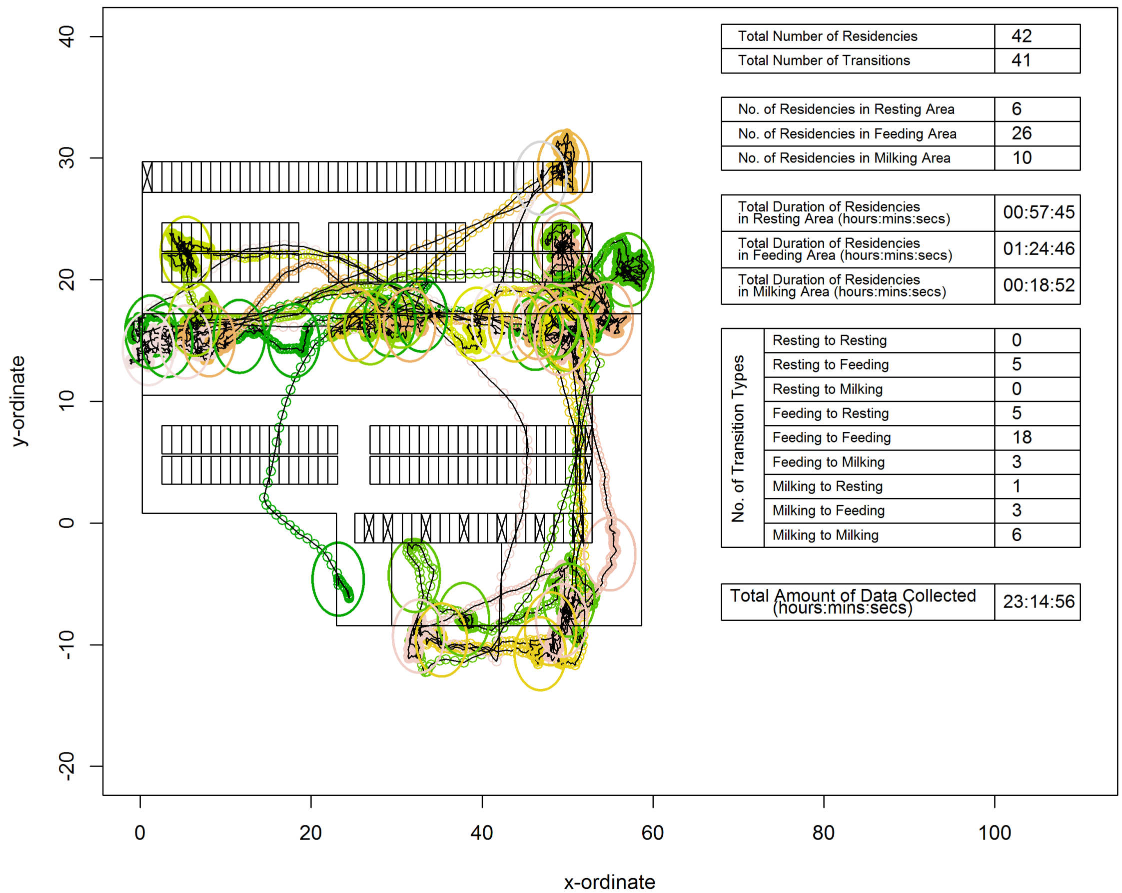
Cow 1892 Day 4 - Out-of-Control Window Size of 11 time points



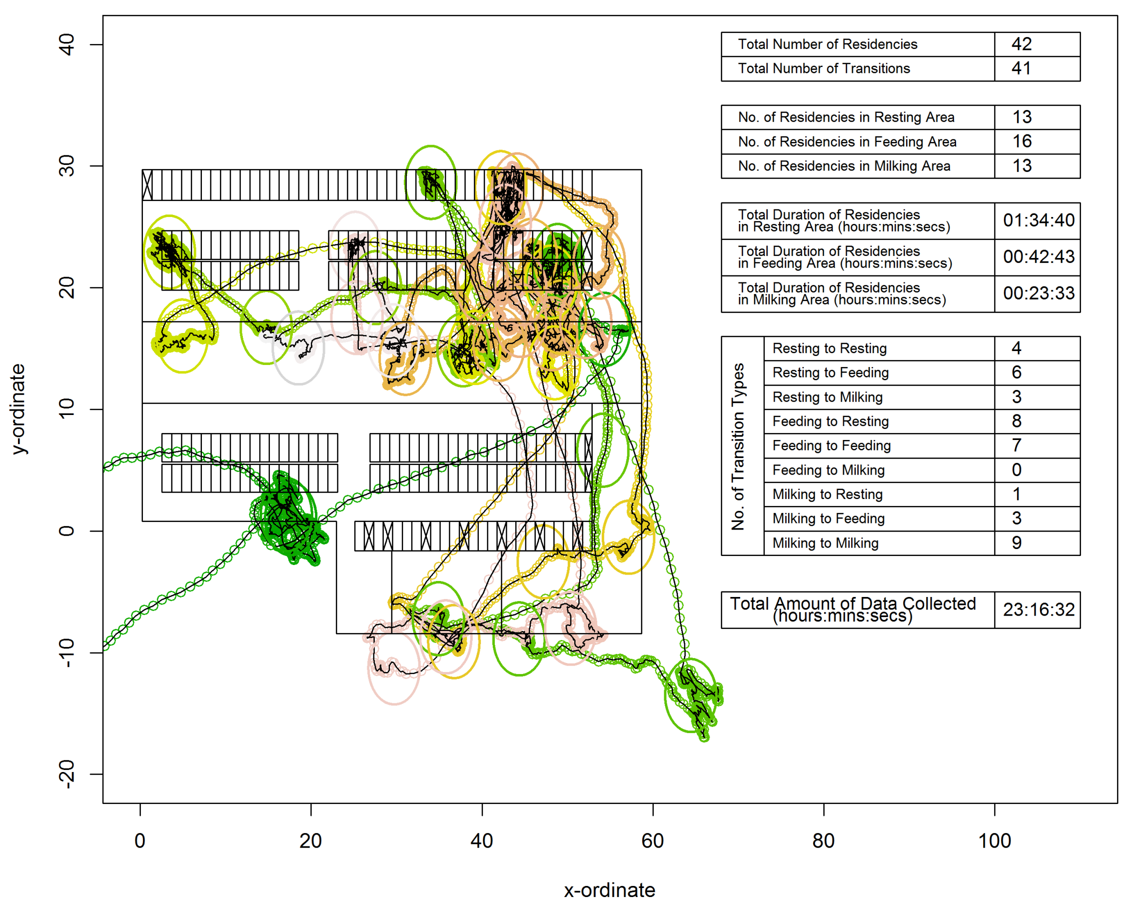
Cow 1892 Day 5 - Out-of-Control Window Size of 11 time points



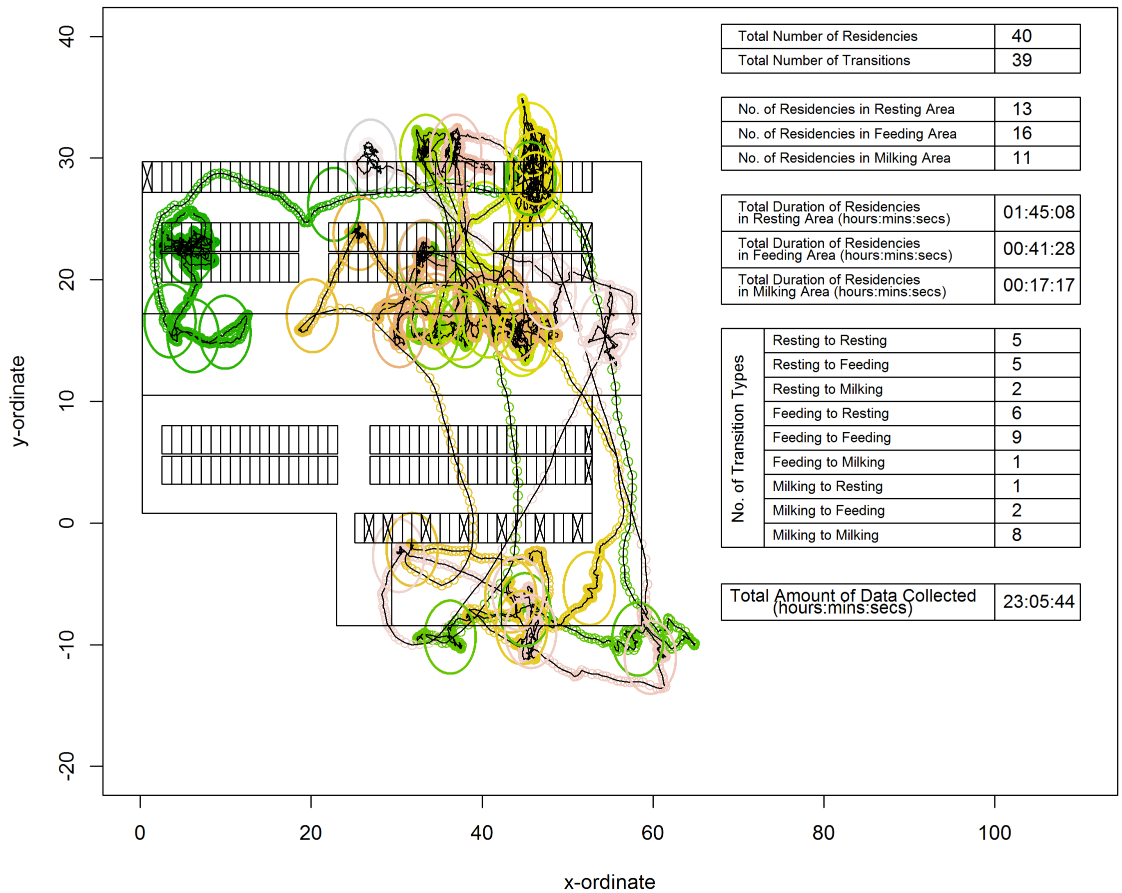
Cow 1892 Day 6 - Out-of-Control Window Size of 11 time points



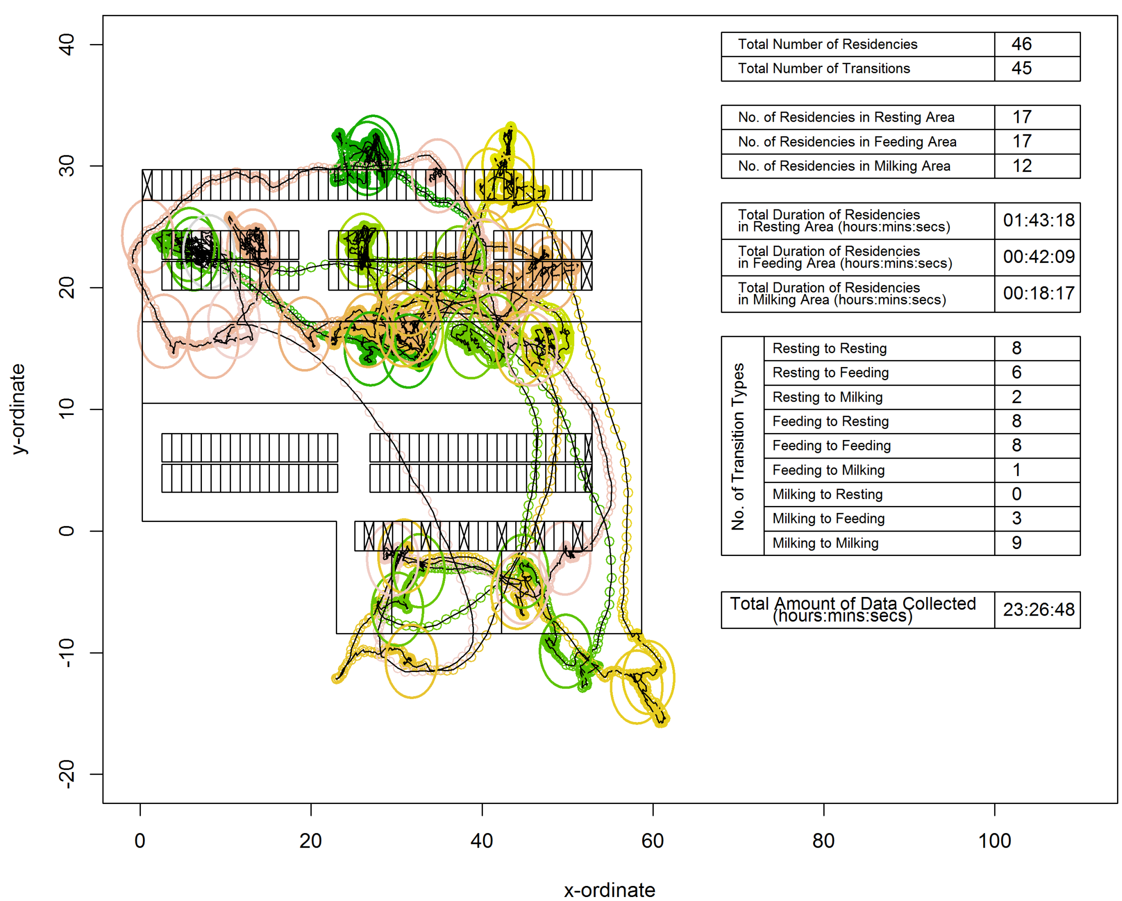
Cow 1491 Day 2 - Out-of-Control Window Size of 11 time points



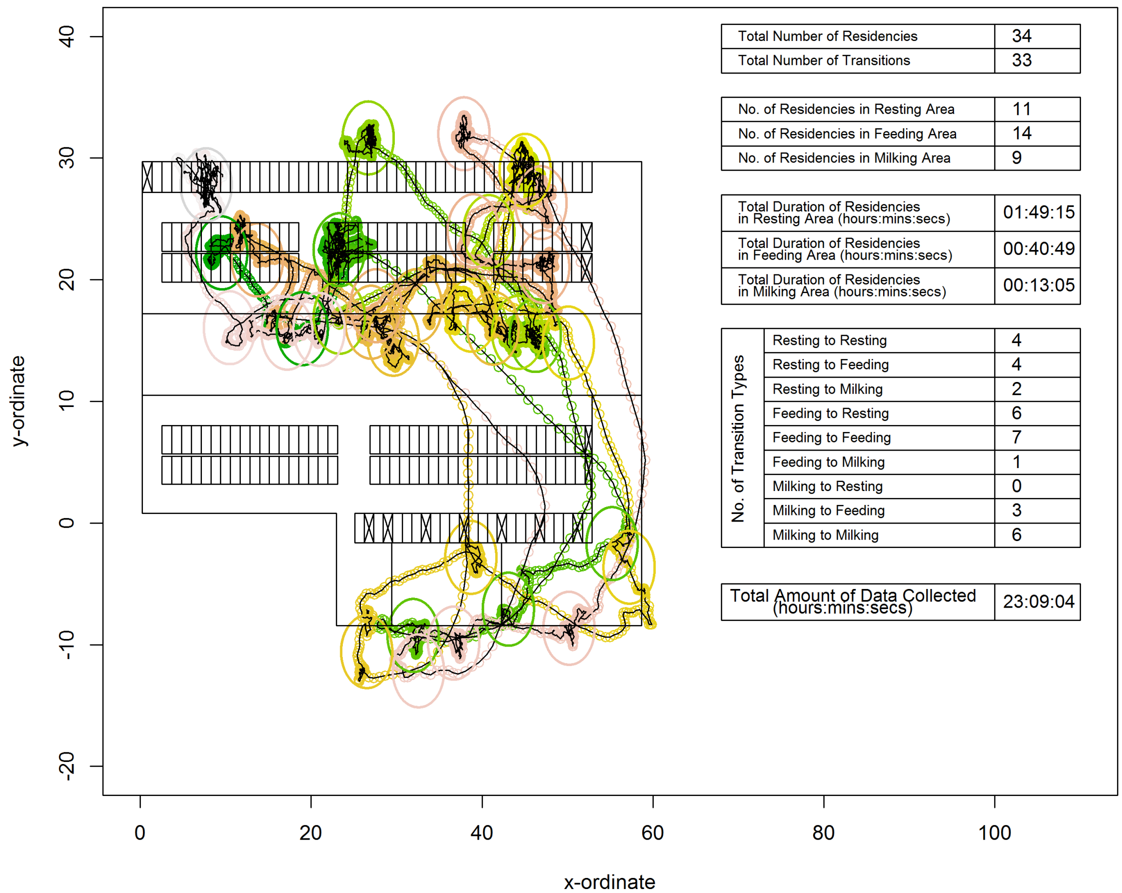
Cow 1491 Day 3 - Out-of-Control Window Size of 11 time points



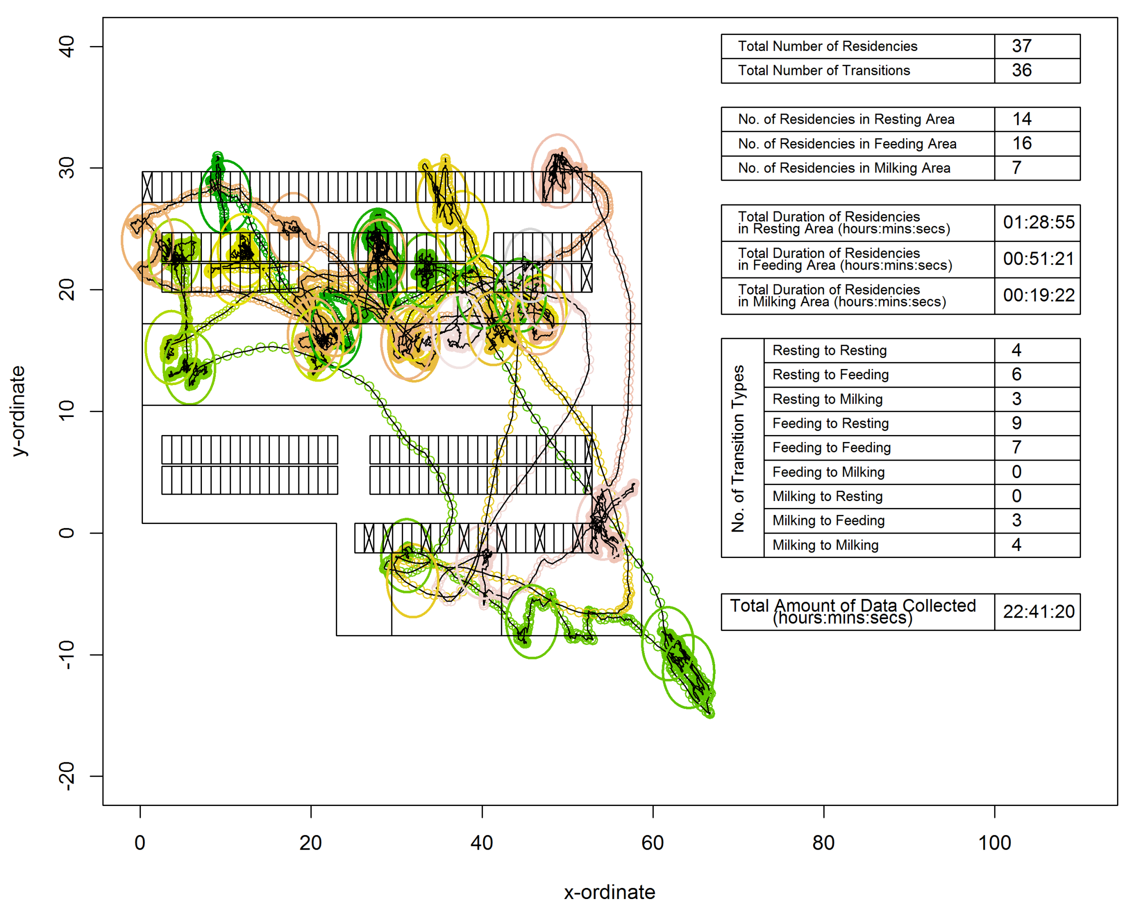
Cow 1491 Day 4 - Out-of-Control Window Size of 11 time points



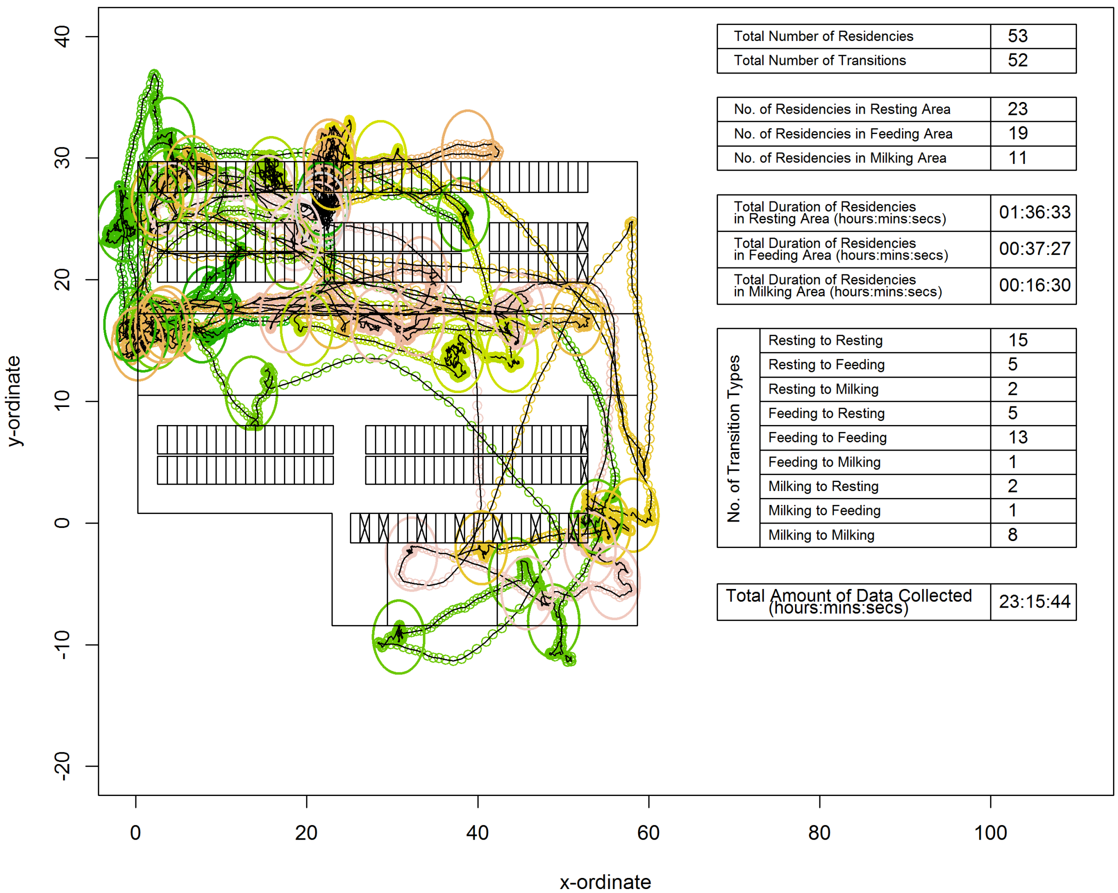
Cow 1491 Day 5 - Out-of-Control Window Size of 11 time points



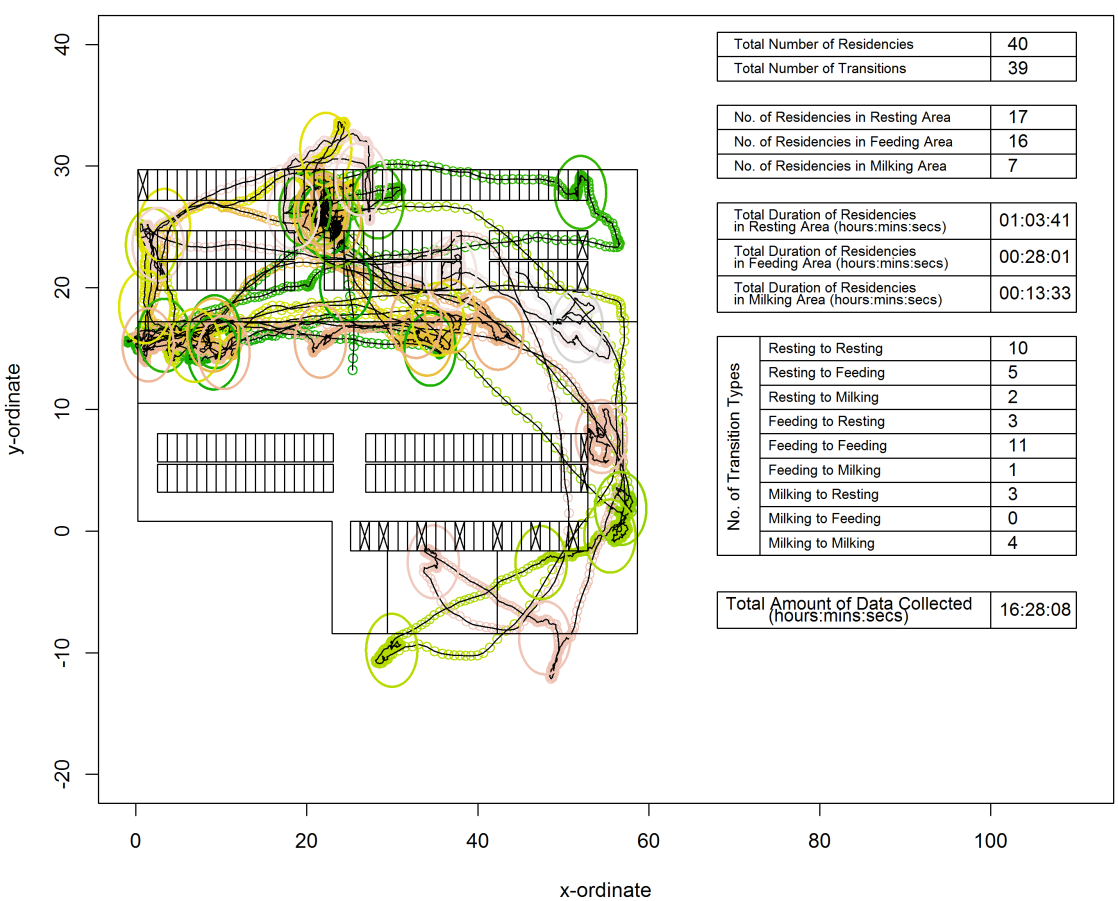
Cow 1491 Day 6 - Out-of-Control Window Size of 11 time points



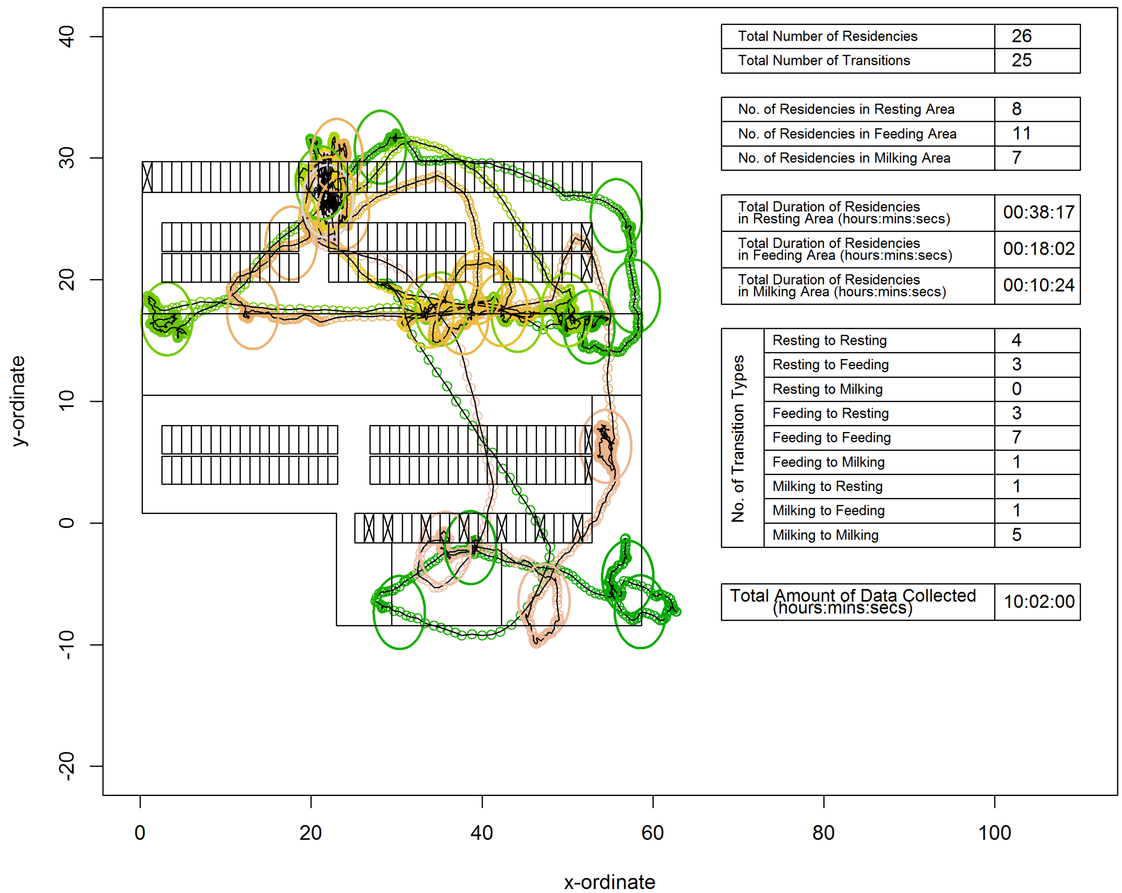
Cow 2596 Day 2 - Out-of-Control Window Size of 11 time points



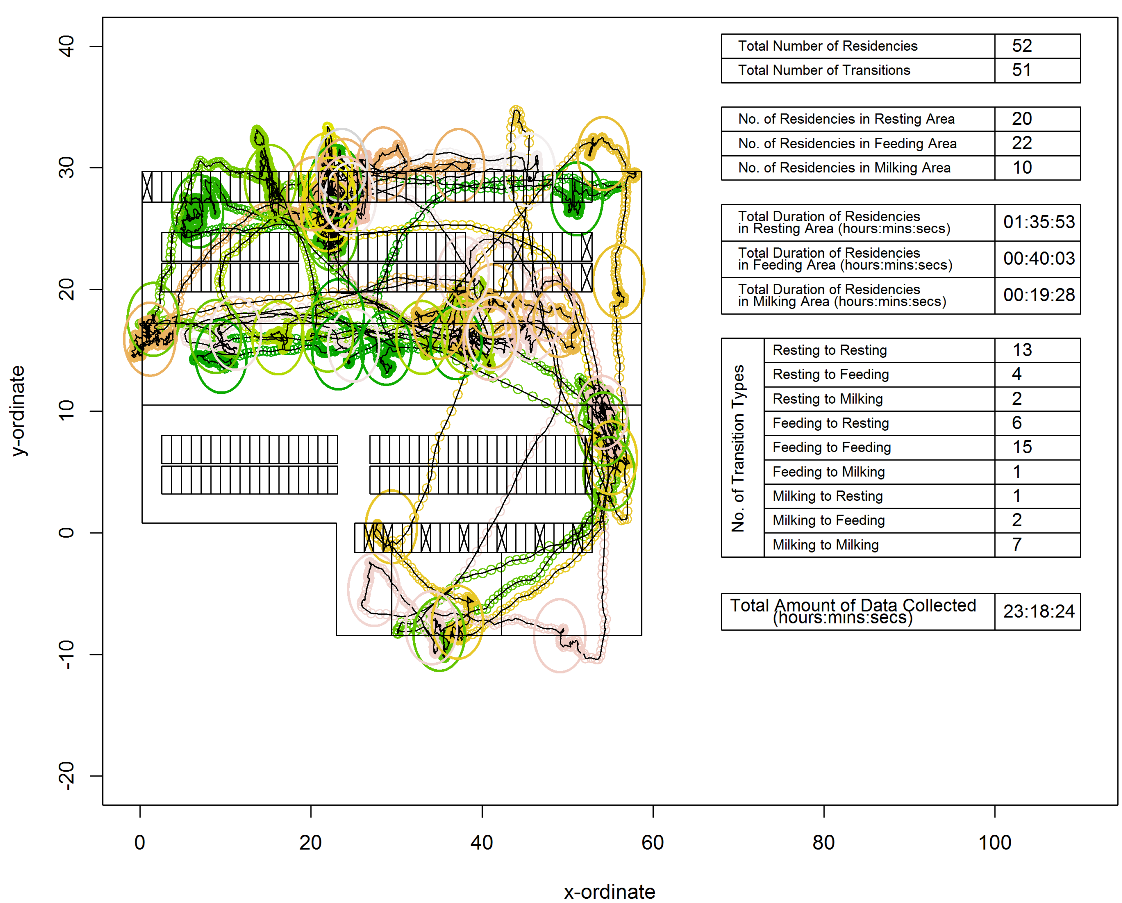
Cow 2596 Day 3 - Out-of-Control Window Size of 11 time points



Cow 2596 Day 4 - Out-of-Control Window Size of 11 time points



Cow 2596 Day 5 - Out-of-Control Window Size of 11 time points



Cow 2596 Day 6 - Out-of-Control Window Size of 11 time points