Technical University of Denmark

DTU

Analysis of data from the MariCare Smartfloor at Skovhuset Care Home

Spooner, Max Peter; Jørgensen, Thomas Martini; Thyregod, Camilla; Ersbøll, Bjarne Kjær

Link back to DTU Orbit

DTU Library
Technical Information Center of Denmark

Analysis of data from the MariCare Smartfloor at Skovhuset Care Home

Report

Prepared by Max Peter Spooner, Thomas Martini Jørgensen, Camilla Thyregod, Bjarne Kjær Ersbøll of DTU Data Analysis

In cooperation with Charlotte Kock Petersen of Skovhuset Care Home and Palle Stevn of MariCare

July 2015

1. Summary

In this project data is analysed from a smartfloor which is installed in an elderly care home. Two lines of investigation are carried out. The first uses "event data" from the smartfloor. This data contains every event (bathroom visits, people entering/exiting the room etc.) registered by the floor over a one year period. Control charts are used to investigate a link between the event data and UTI incidence. A clear predictor for UTI is not found, but the value of control charts in this setting is demonstrated. In the second line of investigation "position data" is analysed. The position data is more limited than the event data in that it is extracted manually using image analysis on the smartfloor user interface program. Using the position data, the trajectories traced by a resident moving about their room are visualised and properties such as direction and speed are investigated. A method is found for comparison of trajectories to determine their degree of similarity and this method can identify unusual trajectories in the dataset.

2. Introduction

This project has been carried out in cooperation with MariCare and Skovhuset Care Home, within the framework of OPALL (Offentlige Private Alliancer). Skovhuset Care Home in Hillerød houses 104 elderly citizens. MariCare smartfloor technology is installed within the floor of each of the 104 apartments. The smartfloor registers the movements of the resident, similarly to how a touch screen device registers the movement of a finger on the screen. The floor is calibrated such that activation of certain areas of the floor triggers the registration of an "event". For example, treading on a floor zone near the bed will cause an "exit bed" event to be registered. In addition, by analysing the imprint made by a person on the floor, the floor senses if a person falls over in the room. A diagram of a typical room in Skovhuset is shown in Figure 1, and includes the main components of the room (entrance with tea-kitchen, combined living and sleeping area, bathroom and balcony). The smartfloor is installed throughout the apartment interior, except for in the bathroom. The events generated by the floor are saved in a database. Staff can play back a video animation showing the movements registered by the floor in any room at any time, and examine reports of the events that have occurred. Some of the events in some of the apartments, in accordance with the wishes of the resident and staff, lead to an immediate alarm notification being sent to the staff, so that assistance may be provided. The purpose of the smartfloor is to help the staff at Skovhuset provide better care for the residents, and according to testimony from staff, the floor is successful in doing so. For example, faster response times to falls are made possible. Less direct intervention of staff in the lives of the residents is required due to the smartfloor, which provides better quality of life. In addition, the staff can investigate possible causes for certain behaviours. If a resident is irritable and uncooperative, the staff may observe in the smartfloor system that the resident was frequently out of bed and restless during the previous night. Reasons for this could be investigated, and symptoms which many residents often underreport such as pain may be uncovered leading to a better diagnosis. The smartfloor provides valuable extra information to the healthcare professionals. However, it is necessary for staff to actively select and view the information that may be relevant on a case by case basis. It will often not be apparent to staff that there is any cause to look into a resident's smartfloor data during some period, even though some valuable insight might be contained in that data. The aim of this project is to investigate how the data generated by the smartfloor can be utilised to an even greater extent than previously. This entails applying techniques that can automatically detect changes in behaviour. The staff can then be alerted when a resident displays behaviour that is out of character, allowing care for that resident to be optimised if necessary. In addition, this may enable even earlier diagnosis of illnesses, which would also greatly improve care and save resources.
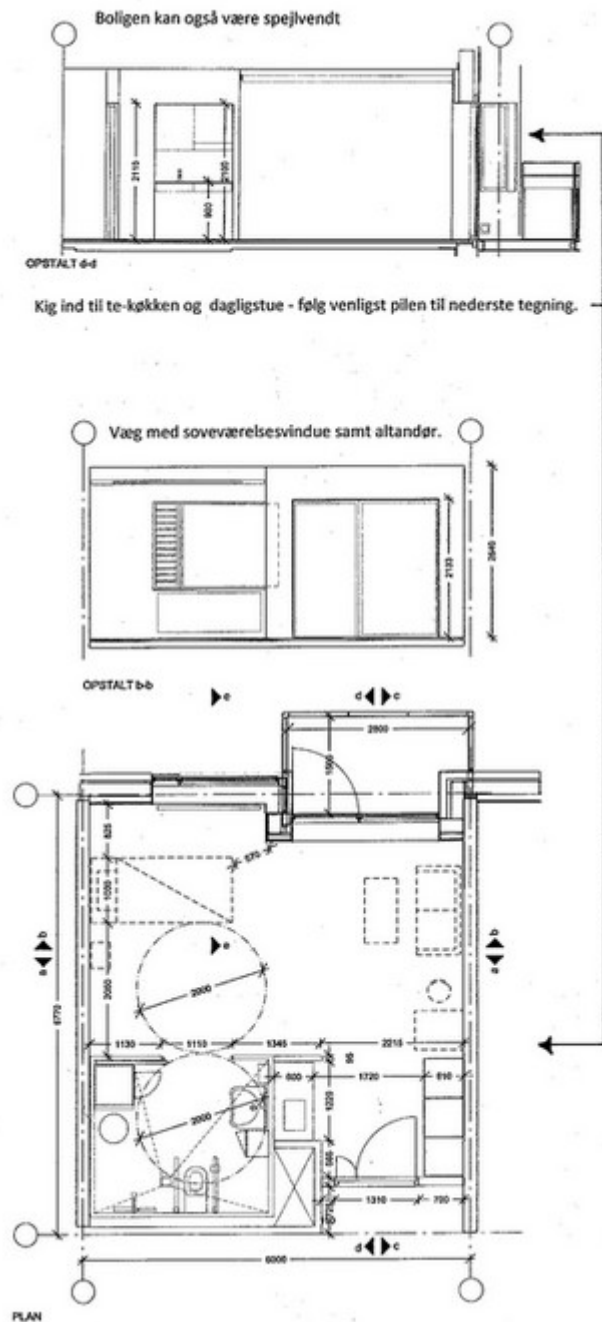
Figure 1 Floor plan of an apartment at Skovhuset (source: http://www.hillerod.dk/da/ForBorgere/Pension_Aeldre/ Plejecentre/Skovhuset/Skitse%20af%20boligen.aspx)

Two different datasets were obtained in this project, each with slightly different aims and different methods of analysis. The first dataset, hereafter referred to as the "event data" consisted of every event registered by the floor in every apartment at Skovhuset over a one year period. This dataset could be accessed directly from the smartfloor database at Skovhuset. The apartment number, type of event, and the exact time of occurrence were contained in the event data. One of the aims for the event data analysis was to investigate whether it could be used to diagnose urinary tract infections. It was hypothesised that individuals would visit the toilet more frequently prior to UTI, and that this can be observed before more obvious symptoms. It is

also thought that individuals would display more restless behaviour, for example during the night, prior to UTI. If these patterns could be identified they can be used for earlier diagnosis of UTI. In the statistical analysis of the event data, control charts are used to investigate these theories. Although a definitive UTI prediction method is not identified, the analysis illustrates how the event data can be utilised to reveal broader patterns of resident behaviour, and this can certainly help to provide better healthcare for the residents.

The second dataset, referred to as the "position data" contains the actual position of a resident over time as they move about in their apartment. This information is saved by the smartfloor system and video animations of the movements can be played back. However the raw data was not directly accessible from the smartfloor system within the scope of this project. To obtain the position data, the video animation of one resident covering 48 hours was recorded and frame by frame image analysis was applied to extract the position and time coordinates of the individual in the room. As this process was quite complex, only position data for a single resident for 48 hours was recovered. This was sufficient data to investigate various techniques. A review of the literature was carried out to find existing methods for analysis of position data, often referred to as trajectory analysis. Using some of these methods, a way to characterise the movements of the resident was found and it was possible to compare trajectories to each other and find unusual trajectories. These promising results show that it should be possible to identify departures from normal behaviour of a resident, simply based on the movement patterns around the apartment, and this can again help with earlier and better diagnosis of problems thereby improving healthcare. However, this would require further analysis using direct access to the position data for more residents over a longer period.

3. Analysis of Event Data

3.1 Specification of Data

The event data was downloaded directly from the smartfloor SQL database at Skovhuset. The database only saves data for 1 year and so the data set covered every event that occurred in Skovhuset between 14/04/2014 and 14/04/2015. The first few rows of the event data are shown in Table 1.

ID	Time	RoomID	Event Type	Zone	Level
2318708	2014-04-14 06:50:52	---	ENTERROOM	-	0
2318709	2014-04-14 06:51:21	---	ENTERWC	-	0
2318710	2014-04-14 06:52:01	---	ENTERWC	-	0
2318711	2014-04-14 06:52:12	---	ENTERROOM	-	0
2318712	2014-04-14 06:52:21	---	ENTERROOM	-	0
2318713	2014-04-14 06:52:28	---	EXITROOM	-	0
2318714	2014-04-14 06:53:30	---	ENTERWC	-	0
2318715	2014-04-14 06:54:03	---	EXITBED	1	0
2318716	2014-04-14 06:54:28	---	GONEBATH	-	0
2318717	2014-04-14 06:55:06	---	EXITBED	1	0
2318718	2014-04-14 06:55:08	---	EXITBED	1	0
2318719	2014-04-14 06:55:30	---	ENTERWC	-	0
2318720	2014-04-14 06:55:43	---	EXITBED	1	0
2318721	2014-04-14 06:56:02	---	ENTERWC	-	0
2318722	2014-04-14 06:57:36	---	ENTERWC	-	0
2318723	2014-04-14 06:58:04	---	ENTERWC	-	0
2318724	2014-04-14 06:58:48	---	BALCONY	-	0
2318725	2014-04-14 06:58:54	---	ENTERWC	-	0
2318726	2014-04-14 07:00:35	---	EXITBED	1	0
2318727	2014-04-14 07:01:10	---	EXITBED	1	0

Table 1 First few rows of the event data that was extracted from the database

The six variables are defined as follows:

ID: A unique ID number for each registered event
Time: Time that the event occurred to nearest second
RoomID: The room in which the event occurred. Note that the RoomID is not the actual apartment number used in Skovhuset. A separate data file showed which Skovhuset apartment number corresponded to each RoomID number. For confidentiality reasons, apartment numbers and RoomIDs are anonymised in this report.
Event Type: The type of event that occurred. There are 10 possible event types as follows
- BALCONY   Someone leaves the apartment via the balcony door
- BALCONYTIME   Someone stays too long on the balcony (30 minutes by default)
- BURGLAR   Someone enters the room unexpectedly from the balcony
- ENTERROOM   Someone enters the room
- ENTERWC   Someone enters the bathroom
- EXITBED   Someone leaves the bed
- EXITROOM   Someone leaves the room
- FALLEN   Someone falls over
- GONEBATH   Someone has stayed too long in the bathroom
- MISSING   No movement in the room for 24 hours

Zone: Is only relevant for events of type EXITBED. If there are several beds in the room, the Zone variable indicates which of the beds triggered the EXITBED event. However, there are only 5 rooms out of the 104 for which zone variable takes on more than one value. For the most part this variable can be disregarded. If it has relevance for particular cases in the analysis it will be noted.
Level: Indicates whether the event triggered an alarm to the staff. Most events are registered in the system but do not result in an alarm. In some cases (chosen by staff together with each resident) an event type is set to give a direct alarm to the staff. This is indicated by the Level variable.

In total, the dataset contained 2,827,965 events that occurred in the building in the one year period. It should be noted that not all event types are registered in all rooms. Residents can choose to have which event types that the floor registers. Also, events can be triggered by anyone inside the rooms and so do not necessarily correspond to the resident who lives there. For example, staff assisting the resident with washing may trigger several ENTERWC events as they go in and out of the bathroom.

3.2 Exploratory Data analysis

The following plot shows the distribution of the total number of occurrences of each event type for all the rooms.
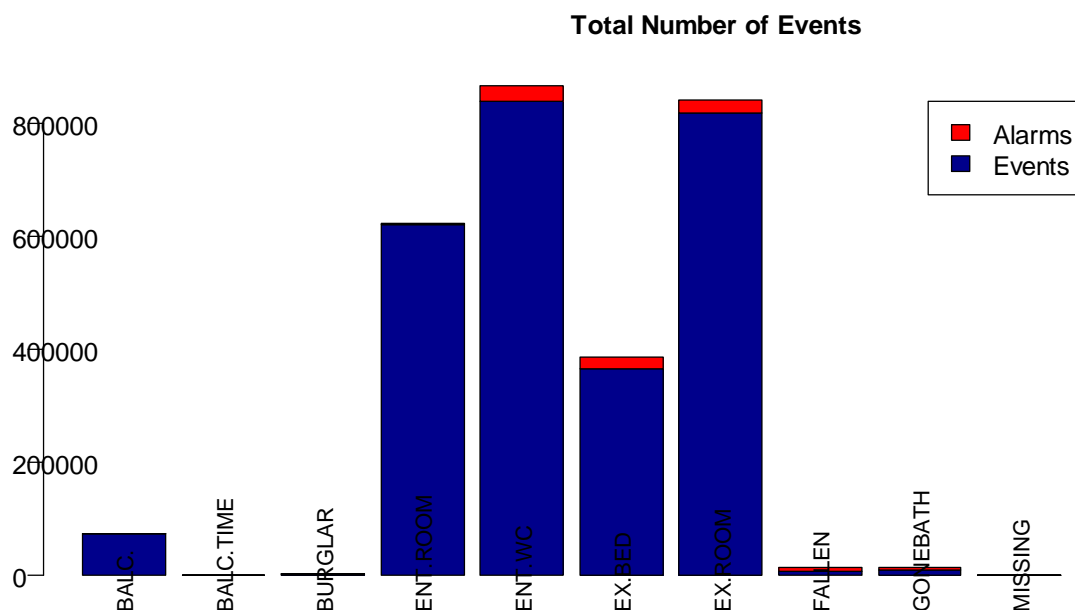

Figure 2 Total number of events for each event type

The ENTERWC and EXITROOM event types occur most frequently, each with over 800,000 occurrences during the year. A distinction can be made between event types that occur often, and those which occur rarely as follows:
Rare Event types: BALCONYTIME, BURGLAR, FALLEN, GONEBATH and MISSING
Frequent event types: ENTERROOM, EXITROOM, ENTERWC, EXITBED, BALCONY

Figure 2 also shows the proportion of events that triggered an alarm. Only a small proportion of the events triggered an alarm.

In Figure 3 the distribution of event types for each month is shown. As the dataset starts halfway through April 2014 and ends halfway through April 2015, there are far fewer events in these months. Apart from this, the relative numbers of event types remains fairly constant throughout the year.
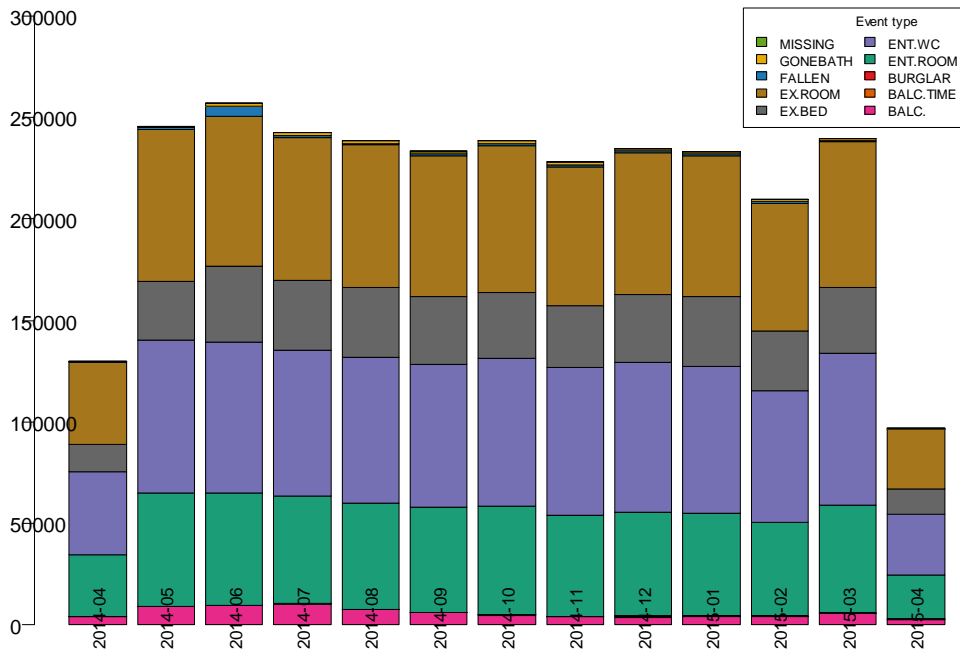
Distribution of event-type ar

Figure 3 Total number of events per month for each event type

In Figure 4, the total number of each event type to occur for each hour of the day is plotted. There is more activity during the daytime hours from around 8 am until 10 pm.
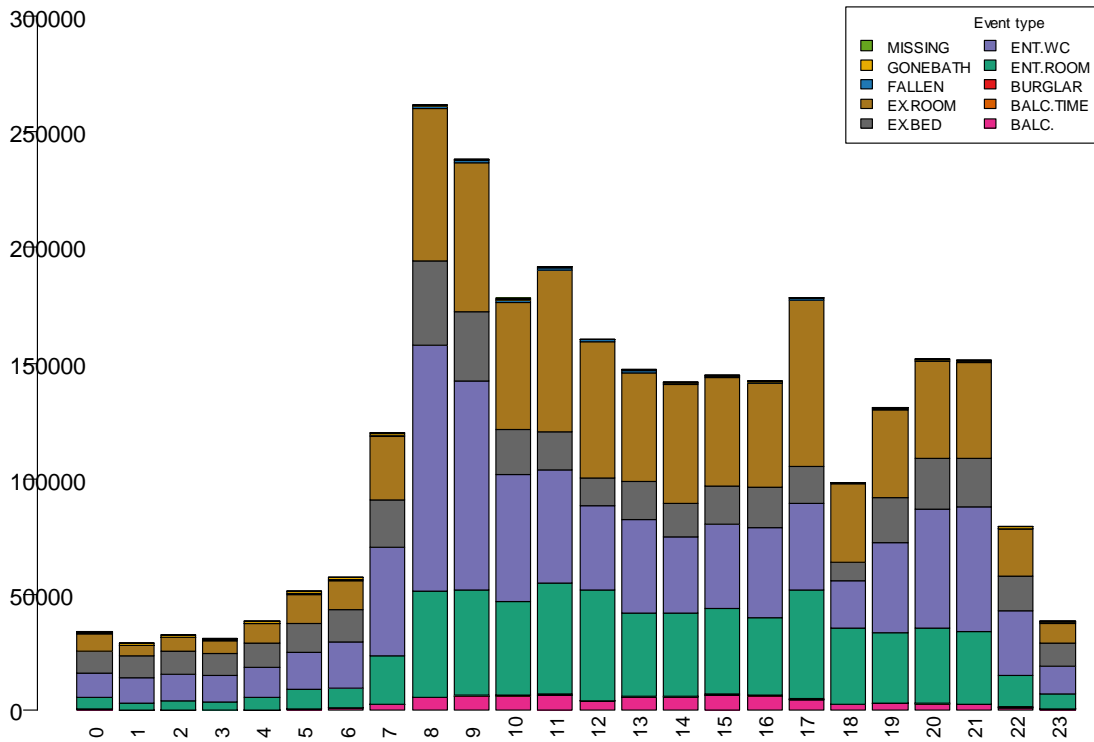
Figure 4 Total number of events for each hour of the day for each event type

Next, the event occurrences can be examined for each room individually. First, the monthly counts of each event type for some randomly chosen rooms are shown in Figure 5. This shows that there is great variation between different rooms. Also, the number of each event type can vary greatly from month to month for the same room.
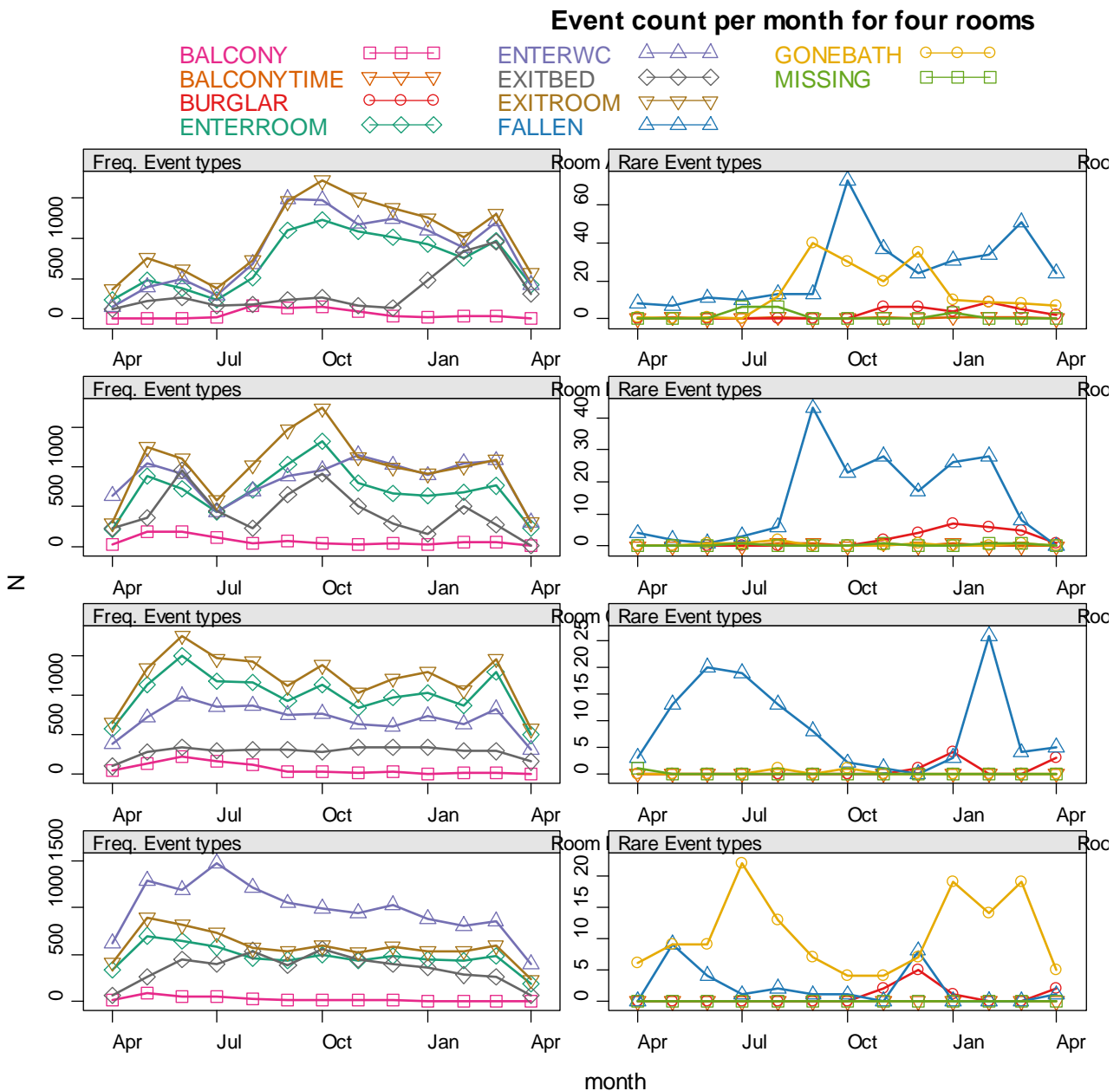
Figure 5 The number of events per month for each event type for selected rooms. Each row is the same room – the left hand side shows the frequent event types and the right hand side shows the rare event types.

For a more detailed view, the daily count of each event type for several randomly chosen rooms is shown. Due to the amount of data, it is not possible to follow the day to day changes in the counts, but the plot does give an overview of the general level of the daily counts over the year. The daily counts for Room A remain fairly stable throughout the year. For Room B, the daily count of ENTERWC appears to be at a lower level in October/November, compared to the rest of the year. Room C appears to be empty around 1. October, as the levels of all event types stay around 0 there. Finally, Room D suggests that the EXITROOM event was deactivated from October onwards.
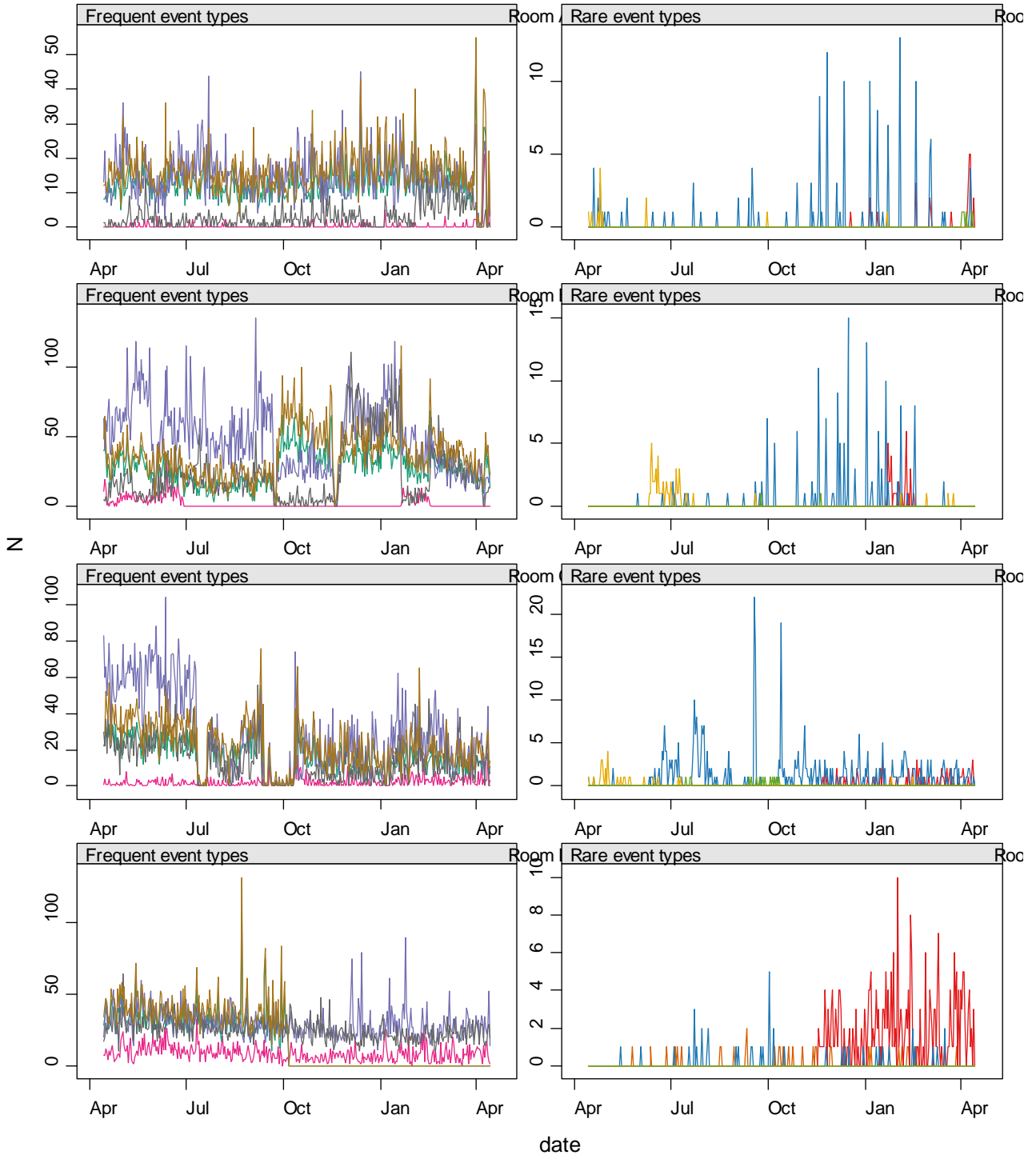
Figure 6 The daily counts for each event type for selected rooms. Each row is the same room – the left hand side shows the frequent event types and the right hand side shows the rare event types.

In Figure 7 an attempt was made to summarise the daily counts of each event type for each room in one place. For each event type, the minimum and maximum daily count for each room is plotted as a black line, with a horizontal black dash showing the mean daily count. If all daily counts were zero for a particular room and event type, then a black "x" is shown in the plot. The shading in the plot indicates which floor the room is on (light blue to dark blue indicating ground floor to 3rd floor respectively), however the order of the rooms within each floor is randomised to ensure anonymity. The plot shows, for example, many rooms can have over 100 EXITROOM events in one day. This seems like a large number, even accounting for coming and going of staff and visitors. It seems likely that not every EXITROOM event truly corresponds to a person exiting the apartment, and similarly for the other event types. Even so, the events should be an indication of the different activities. The summary plots for ENTERWC, EXITBED and FALLEN show that the 79th room in the plot has far greater maximum daily count values for these event types. Perhaps a technical failure occurred in the floor in this room. The daily counts for this room are examined directly in Figure 8 and show that the EXITBED event behaves strangely from June to October 2014, and the FALLEN event is extreme around June 2014, so there is a strong indication that the data for this room is not reliable. It was mentioned at a meeting that one of the rooms experienced water damage to the floor, and probably this is the reason for the unusual event behaviour.
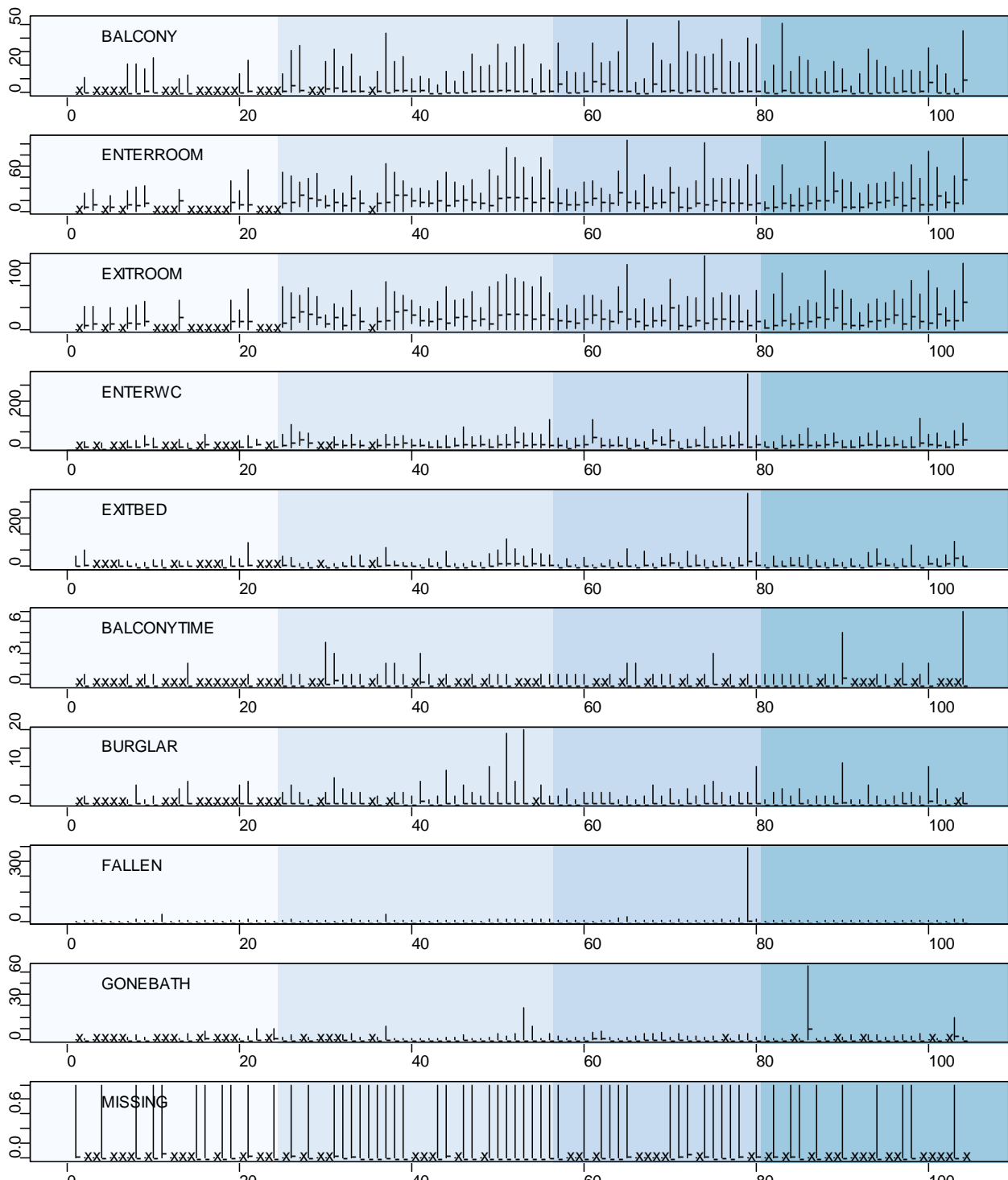
Figure 7 For each event type, and each of the 104 rooms, are black line is drawn from the minimum daily count to the maximum daily count. A dash indicates the mean daily count. If the count was zero every day for an event type and room then an black X is drawn.
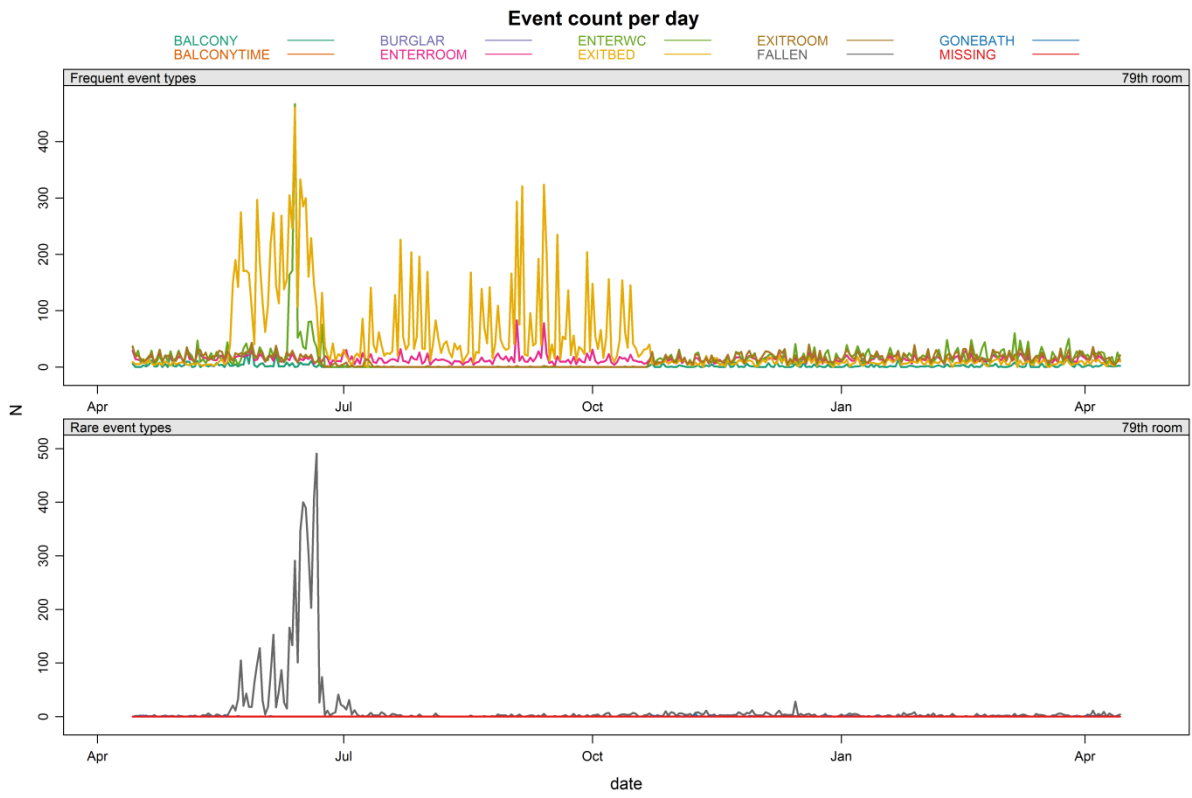
Figure 8 Plot of the daily count for each event type for the room which appears to be an outlier in the previous figure.

After removing this outlier room, histograms are plotted of the daily counts for all rooms for each event type in Figure 9. The histograms show that on most days, the number of EXITBED, EXITROOM, ENTERROOM and ENTERWC events occur less than 50 times. The BALCONY event occurs almost always less than 10 times per day. On most days, there are not any FALLEN events. The maximum number of MISSING events that can occur in one day is 1, because this event type requires 24 hours of non-activity to occur.
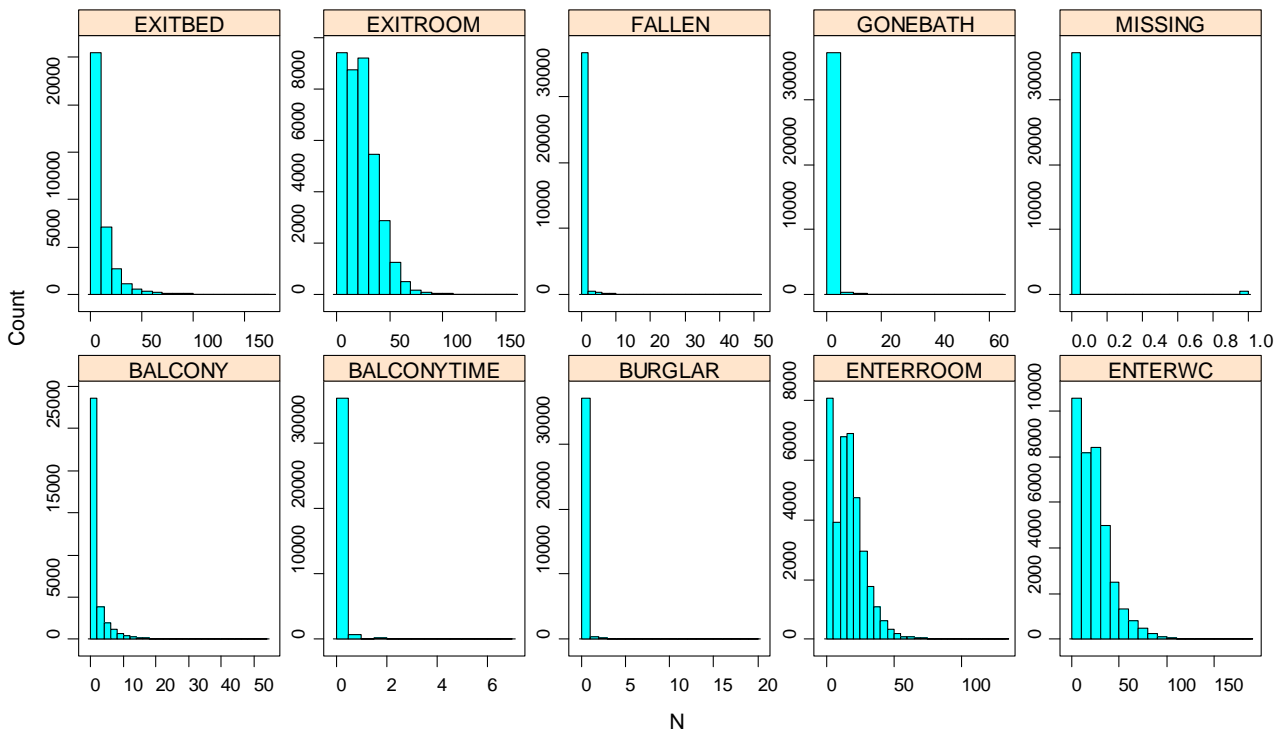
Figure 9 Histograms of the daily counts for each event type for all rooms except the previously identified outlier room.

The final plot for this section is constructed from the average hourly counts for each room and event type. That is, for each room and event type, the number of events per hour is counted, and the average taken for each hour of the day across the 365 days in the dataset. This gives an interesting picture of what an average day looks like for the residents. In Figure 10, these plots are shown for 4 rooms which have a fairly stable event count throughout the year (for example Room A in Figure 6). If the level of the event counts changes greatly for different periods in the year, then it makes less sense to examine the average for the whole year. The variation between individuals can be observed. The resident in Room B in Figure 10 leaves the bed on average around 10am, whilst in Room D the resident often leaves the bed very early around 3 to 4am. In Room C the resident leaves the bed frequently throughout the day. Room B often visits the balcony at 10:00 and at 20:00 whilst Room C visits the balcony around 13:00 and the other two rooms rarely visit the balcony.
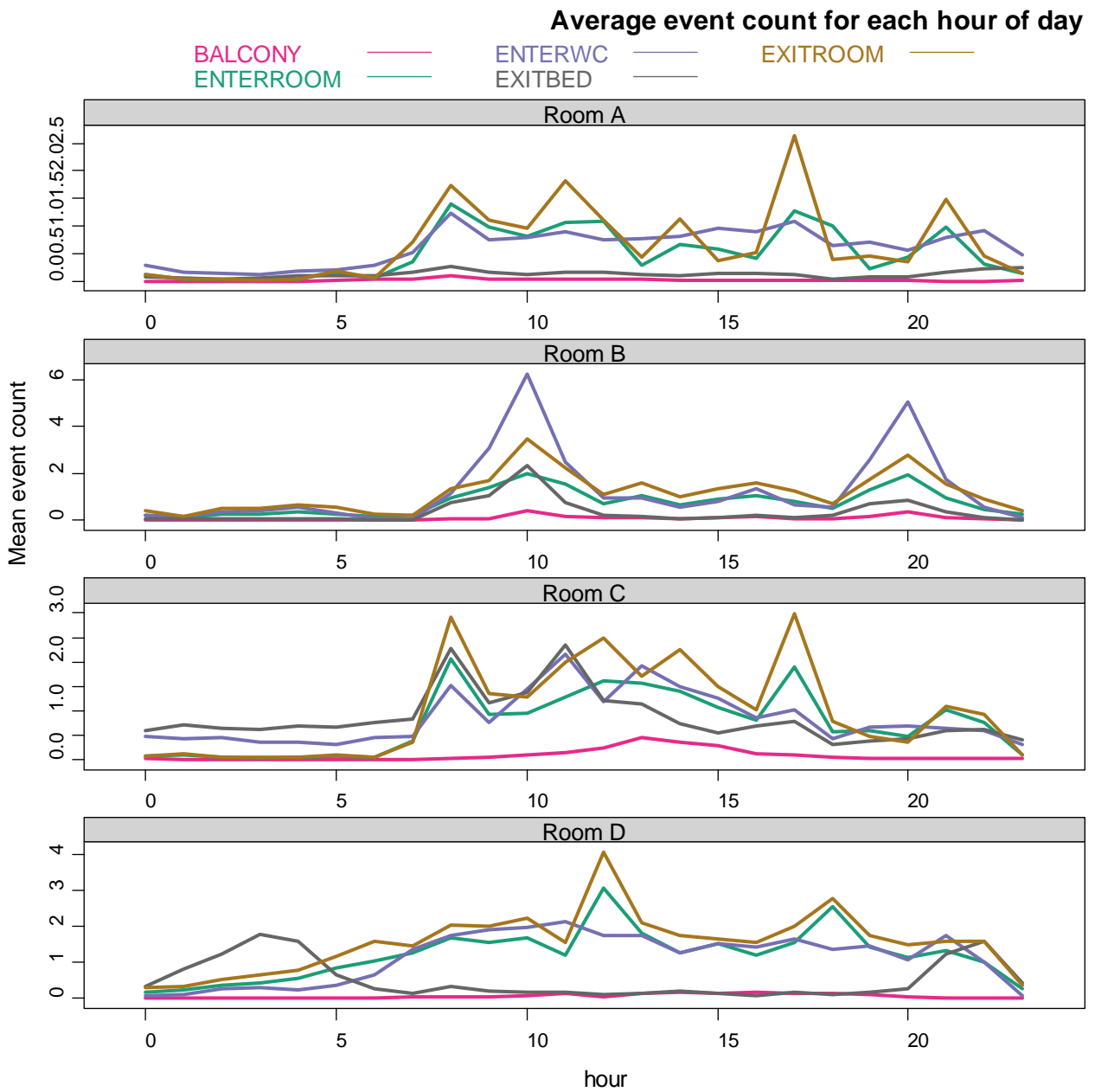
Figure 10 For each hour of the day, the mean number of events to occur during that hour are plotted, for four selected rooms.

3.3   Application of Control Charts

In order to analyse the data with respect to UTIs, background clinical information was provided for three different individuals at Skovhuset which we refer to as Case 1, Case 2 and Case 3.

- Case 1: The resident received treatment for UTI during the periods
    - 22-28/04/2014
    - 20-26/09/2014
    - 17-24/11/2014
    - 12-19/12/2014.
- Case 2: The resident received treatment for UTI during the periods
    - 11-16/09/2014
    - 14-20/01/2015.
- Case 3: The resident received treatment for UTI during the periods
    - 12-18/09/2014
    - 26/11 to 01/12/2014

The room numbers for these individuals is known, and so the event data can be compared with the UTI-treatment periods. It can be assumed that the individuals had UTI some days before the treatments started and probably had symptoms of UTI. The aim is to find some indication in the data, at an earlier date to the commencement of treatment that is correlated with the future onset of UTI. Before discussing control charts, the raw data for these three cases should be examined to check if there are any obvious unusual features in the data. In Figure 11, the daily counts of the events are shown for the three cases, and vertical lines indicate the start and end of the UTI treatment periods.
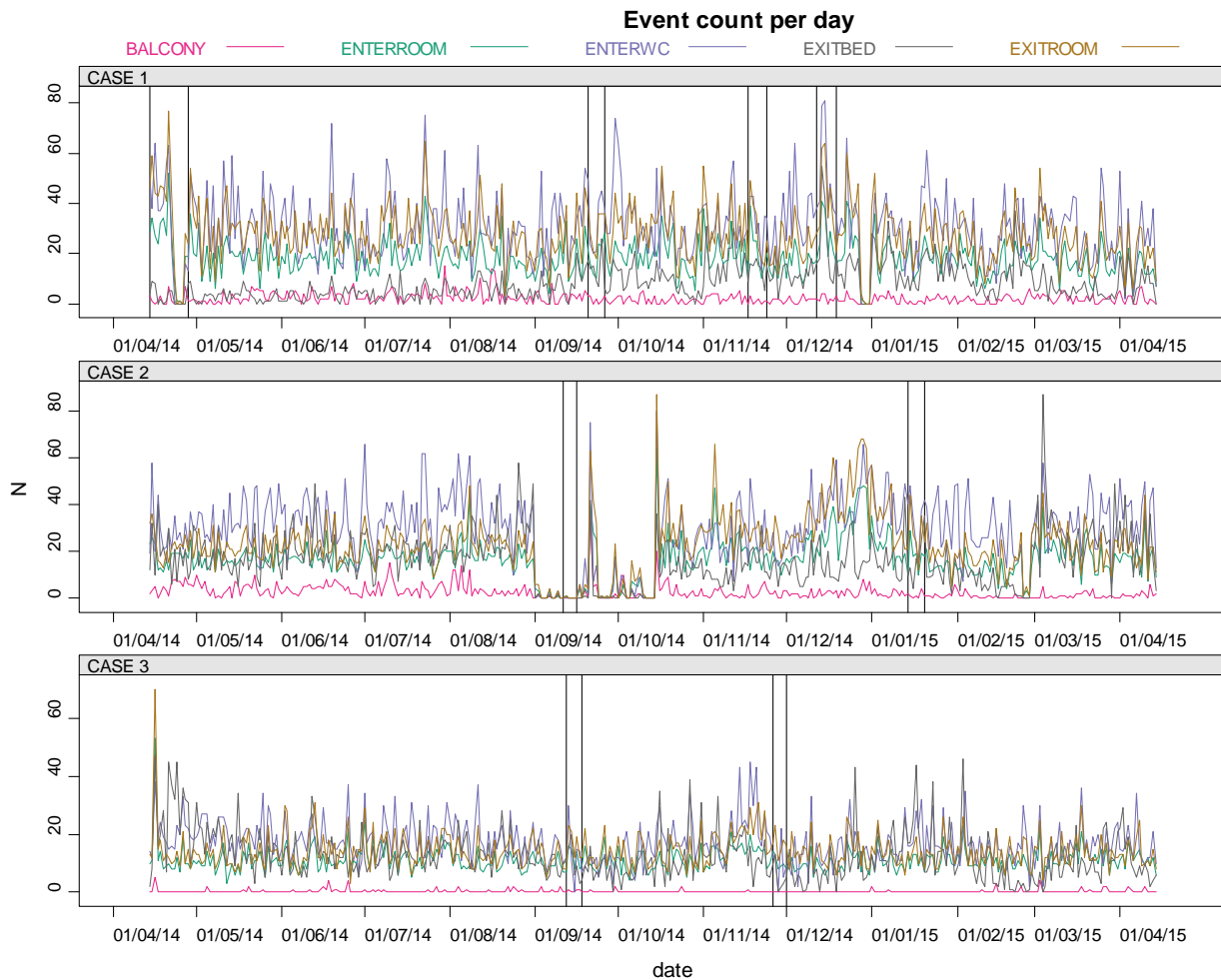
Figure 11 For the three known cases of individuals who were treated for UTI, the daily counts of each event type are shown. The UTI treatment periods are indicated by the vertical black lines.

In case 1, the first treatment period is at the very start of the available data, and so it will not be possible to predict the first UTI occurrence due to absence of any prior data. In case 2, there appears to be extremely low activity in the room in the first half of September 2014, suggesting the room is empty, so again it will not be possible to predict the UTI incident that started on 11. September. Otherwise, there is no obvious signal prior to the UTI periods in Figure 11.

The control chart is a statistical tool devised by Walter A. Shewart in the 1920s. They are widely used to detect changes in industrial manufacturing processes but are also seeing increasing applications in healthcare (Woodall, 2006). In the present setting, instead of detecting an atypical change in the attributes of some manufactured items, we are interested in detecting an unusual change in the event occurrences generated by an individual. Of course, we expect day to day changes in the particular events registered by the floor, as people do not behave exactly the same way every day. The aim is to identify a greater than usual change in the data. Control charts were developed precisely to meet this goal of detecting when variation becomes greater than what would usually be expected.

Control charts are used to monitor a process. A quality characteristic for the process is chosen. An example process would be the production of electronic printed circuit boards, and the quality characteristic could be the number of defects on the boards. A sample of 100 circuit boards is taken at regular intervals from the

process, and the number of defects is plotted. The mean of the sample statistic is calculated and added to the plot as the center line. Upper and lower control limits are added to the plot typically 3 standard errors from the center line. Points that lie outside the upper and lower control limits are then indications that there is more variation than usual in the process, and the process is said to be out of control. These principles are illustrated in Figure 12.
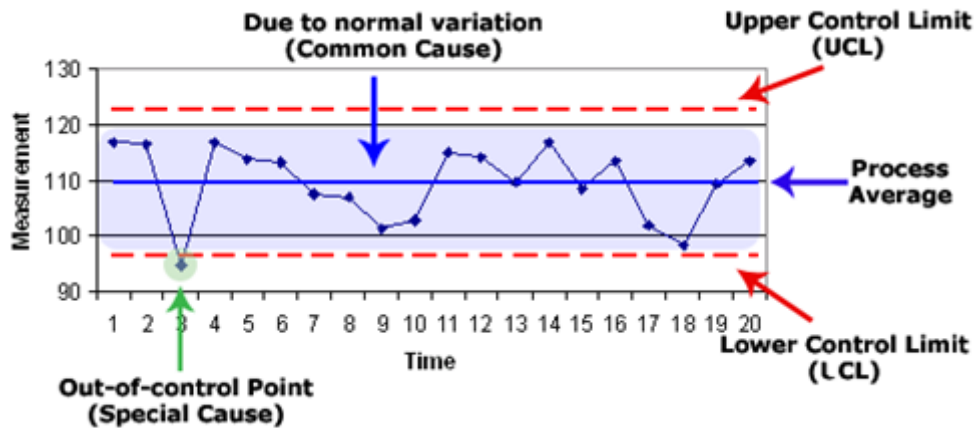

Figure 12 Example illustrating the main components of a control chart.

For the event data, it would be logical to monitor, for example, the number of ENTERWC events that occur each day for an individual. An early symptom of UTI is more frequent visits to the toilet. Control charts offer a solution to identifying what is the usual number of toilet visits for a person, and when the number of toilet visits increases beyond what would usually be expected. The appropriate type of control chart for count data is known as a C-chart, and the upper and lower control limits are usually calculated using the formula $c \pm 3\sqrt{c}$, where $c = \bar{x}$. This is based on the assumption that the variable is Poisson distributed. However, the histogram of the ENTERWC daily count (Figure 13) shows that this variable is not well represented by a Poisson distribution. In the Poisson distribution, the variance is equal to the mean. It appears that in the ENTERWC counts, the variance is larger than the mean, and so the Poisson distribution (black line in Figure 13) does not fit the data very well. If the usual control limits are used, too many out of control points will be observed. One way to allow for the mean and variance to have different magnitudes is to use a "generalised Poisson distribution". The generalised Poisson distribution has two parameters. These are estimated from the data using the method of moments, and the resulting probability mass function is plotted in red in the histogram. For details of the generalised Poisson distribution probability mass function and parameter estimation see Xie & Thong, 2006.
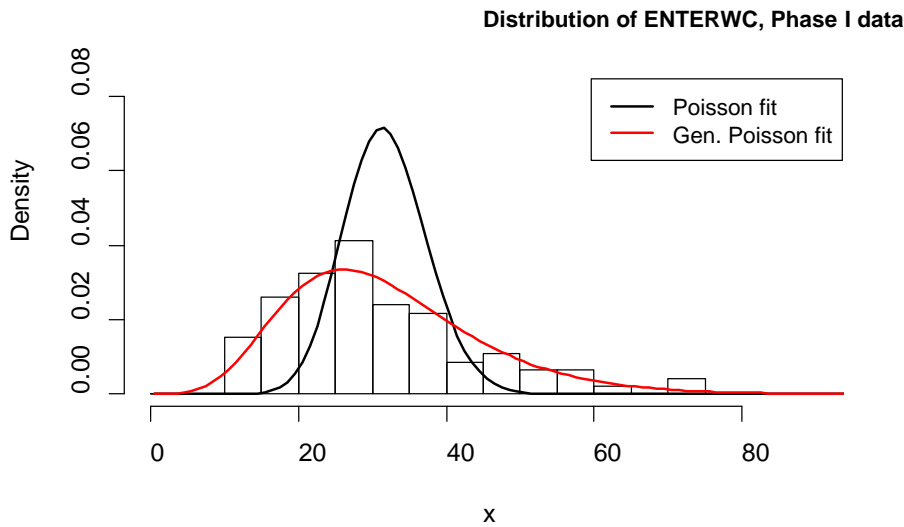
Figure 13 Histogram of the ENTERWC daily count for the case 1 Phase I data. The closest fit using the Poisson distribution is shown in black, and the closest fit using the Generalised Poisson distribution is shown in red.

Once the parameters of the generalised Poisson are found for the data, the control limits are calculated using the formula

$$Center\ Line = \bar{x}, \qquad LCL = \max\{x \mid \Pr(X < x) < 0.005\}, \qquad UCL = \min\{x \mid \Pr(X > x) < 0.005\}$$

This means that when the process is in control, roughly 1% of the points are expected to fall outside of the control limits.

The control limits should be calculated based on data from a period when the process is thought to be in control (known as Phase I data). For case 1, it is supposed that the process is in control for the period 1. May to 31. July 2014, as there are no UTI treatments in this period. Using the methods just outlined, the control chart shown in Figure 14 was constructed. The Phase I period used to determine the control limits is shaded in green. UTI treatment periods are shaded red. In addition to indicating the points that fall outside the control limits with red, two additional rules are applied:

- Rule 2: If 9 points in a row are either all above the center line, or all below the center line, then the 9th point breaks Rule 2 and is coloured blue in Figure 14.
- Rule 3: If 6 points in a row are continually increasing, or continually decreasing, then the 6th point breaks rule 3 and is shaded orange.

In the week preceding each UTI treatment period, there are no out of control points in the control chart, and so this control chart would not assist in earlier diagnosis of UTI in this particular case. During the last UTI treatment period in December, there are two points above the upper control limit. In addition, starting 4 days prior to the treatment period and lasting until the last day of the treatment period, all of the points lie above the center line, leading to the final 3 points breaking rule 2. This suggests an increase in the level of the ENTERWC variable. However, this increase is not captured by the control chart prior to the treatment period, and so would not aid in earlier diagnosis.

The daily count of ENTERWC did not result in a clear warning signal of UTI in this case. Therefore, consider the count of ENTERWC events that occurred during the night time only for each date. I.e. the number of ENTERWC events occurring in the period 22:00 to 08:00 and call this variable ENTERWC-night. The count of ENTERWC events occurring during 08:00 to 22:00 of each date is also made into the variable ENTERWC-day. The control charts for these two new variables are constructed using the same methods as above and shown in Figure 15.
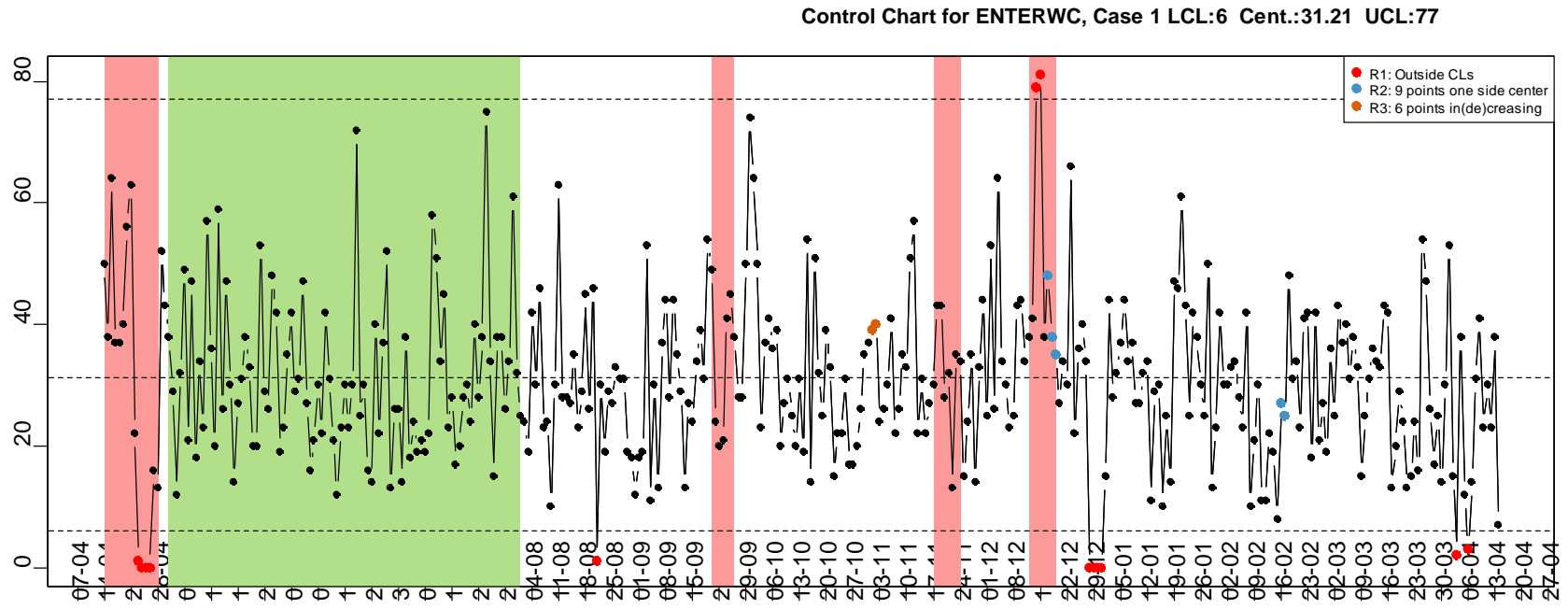
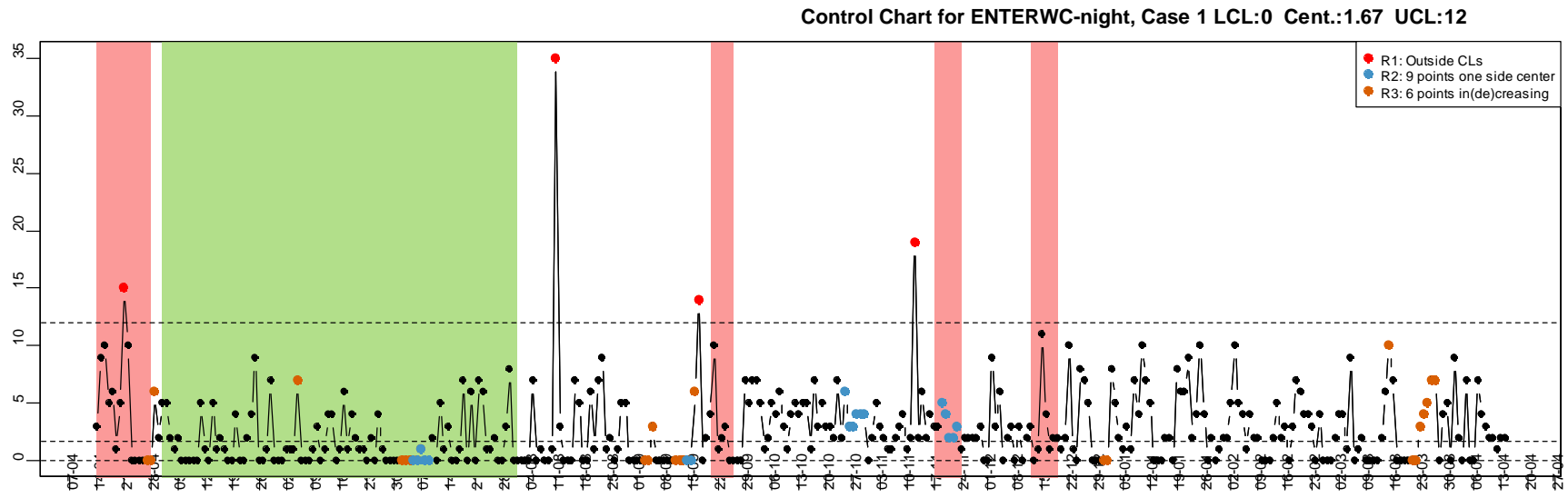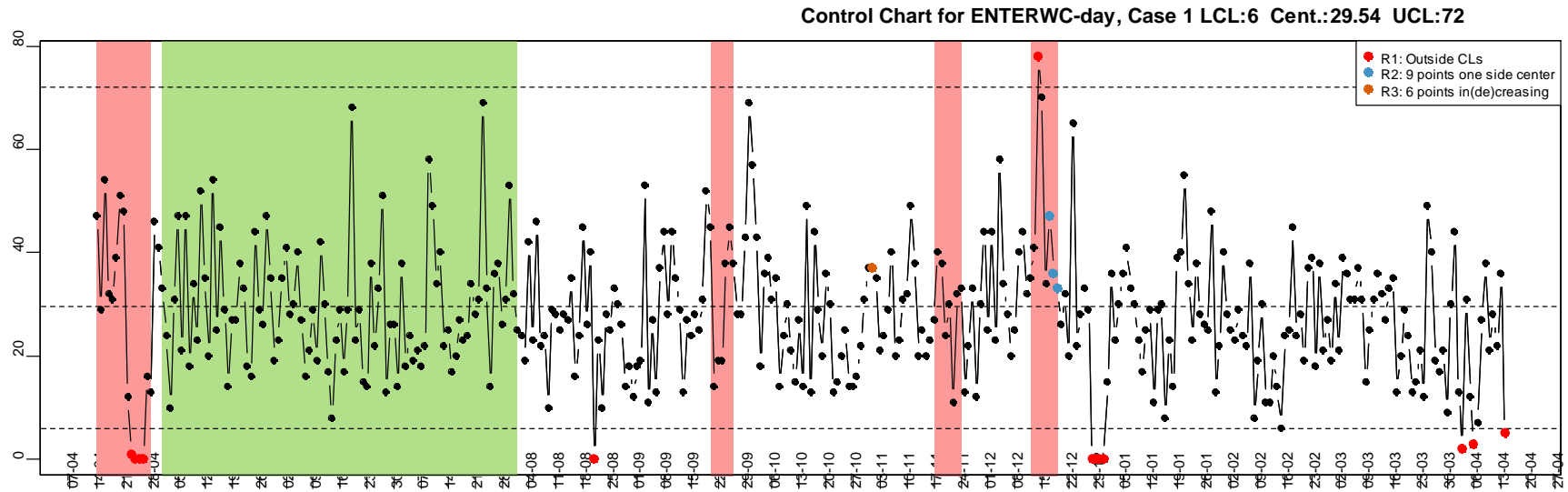**Figure 14 Control chart for case 1, ENTERWC daily count**

Figure 15 Control charts for the new variables ENTERWC-day (top) and ENTERWC-night (bottom) for case 1

The behaviour of the ENTERWC-day variable appears to be fairly similar throughout the year with very few points outside the control limits. However, for the ENTERWC-night variable, it is interesting that a point lies above the upper control limit 3 days prior to the UTI period in September, and also 5 days prior to the UTI period in November. This is consistent with the hypothesis that night time visits to the toilet become more frequent with UTI. It must also be noted that there is an extreme point above the upper control limit on 11. August where there is no recorded UTI incident.

Next, a similar approach is applied to the EXITBED event type. Control charts are constructed for the daily count of EXITBED, as well as the count occurring during the daytime and during the night time of each date and are shown in Figure 16. These charts show that there is an increase in the level of EXITBED events from around 15. September as most points after this date are above the center line, and many points lie outside the upper control limit. From the daytime and night time control charts, it can be seen that the increase in level is entirely due to an increase in EXITBED events during the day time. This pattern indicates a change in the daytime sleeping habits of the individual. The change in level in the EXITBED and EXITBED-day control charts may have assisted in early diagnosis of the September UTI incident, but would not have helped in the two subsequent incidents. The EXITBED-night control chart shows out of control points prior to the September and the November UTI incidents. A correlation can be expected between the EXITBED-night variable and the ENTERWC-night variable, as the most likely reason for leaving the bed during the night is to go to the toilet. This can be observed by comparison of Figure 15 and Figure 16, so in practice it may be best to only use one of these charts.
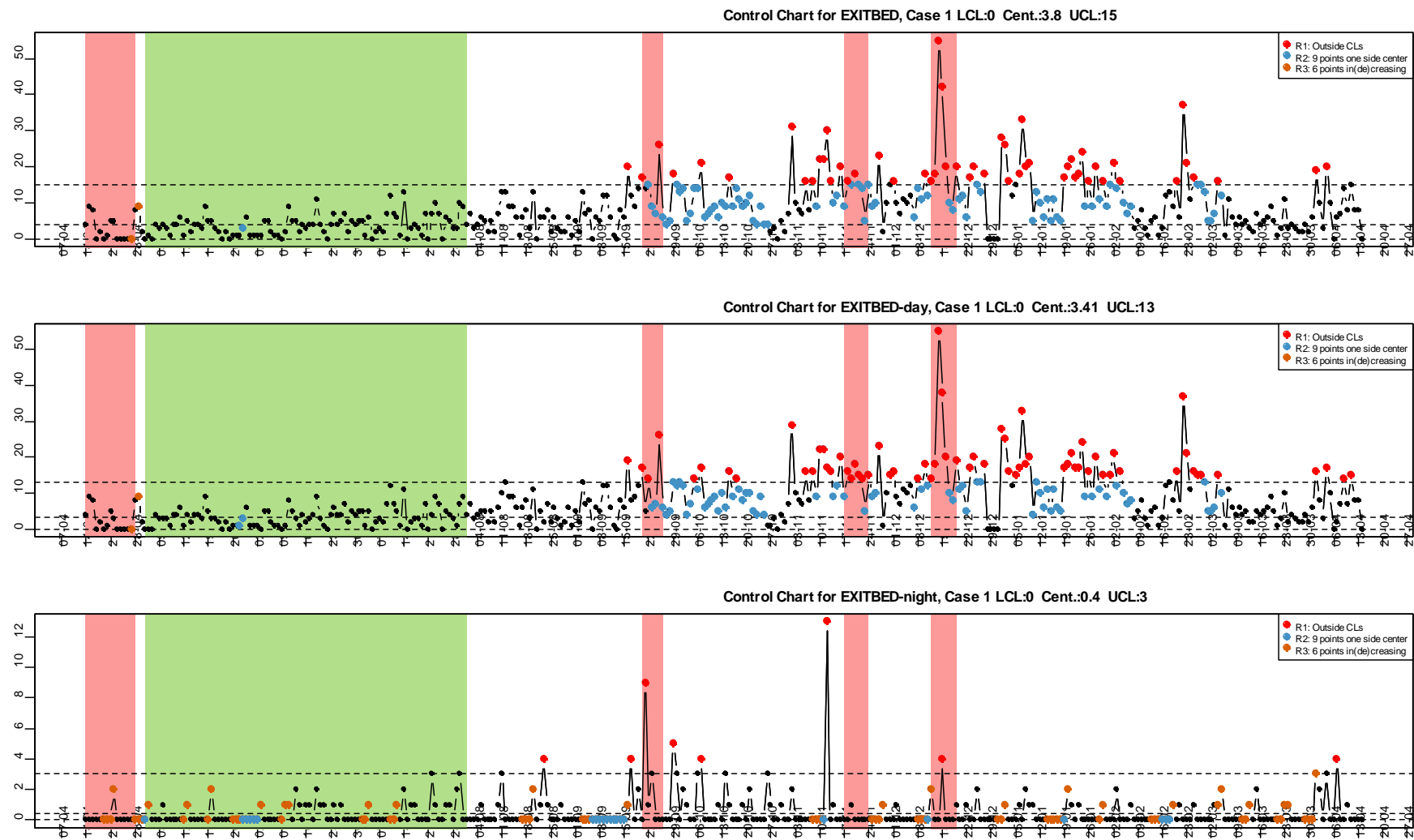
Figure 16 Control charts for the EXITBED daily count (top), as well as the EXITBED-day (middle) and EXITBED-night (bottom) variables for case 1.

The same analysis performed on Case 1 above was also performed on Case 2 and Case 3, and the resulting control charts are included in the appendix. Recall that in Case 2 there is only the possibility of predicting 1 UTI incident as lack of data means the first UTI incident can be disregarded. In Table 2 a brief summary of any points of interest observed in the control charts before the UTI incidents is presented. There were six opportunities for predicting UTI incidents (excluding the first UTI in Case 1 for which prior data was not available, and the first UTI of case 2 for which it appears that the apartment was uninhabited in the days leading up to the UTI). Therefore, the performance of the control charts can be assessed using the success rate of predicting these UTIs.

	Case 1 (3 UTI Treatments)	Case 2 (1 UTI Treatment)	Case 3 (2 UTI Treatments)	Score out of 6
ENTERWC	None	None	Out of control points 6-8 days prior to 2nd UTI	1
ENTERWC-day	None	None	Out of control points 6-8 days prior to 2nd UTI	1
ENTERWC-night	Out of control prior to 2 of the UTIs	None	None	1
EXITBED	Increase in level just before Sept. UTI.	Out of control just before UTI	None	2
EXITBED-day	Increase in level just before Sept. UTI.	None	None	0
EXITBED-night	Out of control prior to 2 out of 3 UTIs	Out of control just before UTI	None	3

Table 2 Summary of the results of the control chart analysis of the three cases

For the three cases control charts were also constructed for the other frequent event types (ENTERROOM, EXITROOM, BALCONY), but are not included here as no obvious pattern was observed. Also, different types of control chart techniques were briefly investigated such as exponentially weighted moving average (EWMA), cumulative sum (CUSUM) and multivariate versions but no discoveries of great importance were made and a detailed description of these results are omitted.

The control chart analysis in this section did not discover a definitive way to anticipate UTI incidents. Even so, they illustrate how changes in behaviour of the individuals can be observed. There is probably a lot of variation between individuals in how they respond to UTI, and so a much larger number of cases would need to be investigated to find a method that can anticipate UTIs in a large proportion of cases.

4. Analysis of Position Data

In this section, work relating to the position data is presented.

4.1 Examples of trajectory analysis in the literature

A review of the literature was carried out to understand what possibilities existed for analysis of position data. A relevant search term was found to be *trajectory analysis,* where a trajectory is the path traced in space by the movement of some object over time. Many different applications of trajectory analysis were found, and only a few examples will be given here.

In Majecka, 2009, the trajectories of pedestrians walking through an atrium in a university are analysed. As with many of the other sources found, much of the report focuses on extracting the trajectory data from video recordings. Once the trajectories (position and time coordinates of the pedestrians) are extracted, methods are investigated to identify "normal" trajectories and "abnormal" trajectories. The problem of how to compare trajectories of different lengths is adressed by fitting a cubic spline to each trajectory, and selecting 11 control points on the cubic spline to represent that trajectory. A Gaussian mixture model is then applied to cluster the trajectories into similar groups, and a trajectory's degree of abnormality is assessed based on its probability according to the Gaussian mixture model.

In Suzuki et al, 2007, the trajectories of the movement of customers in a shop are investigated. The trajectory data is obtained from video anlayis of real surveillance footage of the customers. Hidden Markov Models and K-means clustering is applied to identify different types of trajectories, such as a typical "purchase pattern" or a typical "magazine browsing" trajectory. Abnormal trajectories are identified as, for example, those which do not pass the cash register (possible shop lifters).

In Stauffer & Grimson, 2000, the trajectories of traffic (vehicles and pedestrians) on a road is analysed. Most of their work focuses on extracting the trajectory data from the video footage. To classify the extracted trajectories an advanced method involving generating a codebook of prototype trajectories is described. The classification method could distinguish between people and vehicles, and could be used to detect unusual events.

In Gellert & Vintan, 2006, the movements of employees in an office building are examined. Here, the data does not consist of trajectories, but rather of sequences of rooms/offices visited by an employee. These room sequences are obtained from real persons from a building with technology installed referred to as "smart doorplates". The aim of the authors is to predict the next room a person will visit based on their past history of rooms visited. From the application of Hidden Markov Models, they successfully predict the next room in up to 92% of cases. It is of interest that in this paper, the movements of people are not represented by complicated trajectory data of 3 dimensions (x, y, time), but rather by a simple sequence of rooms visited, and this could be a useful way to reduce trajectory data. Also, these discrete sequences lend themselves well to analyis with Markov Models.

These examples motivated the approach taken to the position data from Skovhuset. A common theme in the examples is to identify movement patterns which are unusual, and this is the same aim adopted in this project. If the smartfloor data in Skovhuset can be used to identify when the movements of a resident become out of character, there may be opportunies for staff to help the resident if necessary. For example, the onset or worsening of dementia may be detected as more frequent and erratic movements around the room at night. Or, if a resident becomes less active, this may be a sign of reduced mobility or depression. The information contained in the movement of the residents could therefore be a valuable diagnostic tool for the staff at Skovhuset. However, they cannot be expected to examine the Smartfloor animation of the residents for every resident 24 hours a day, and so a method is needed which can analyse the movement data and identify unusual patterns automatically. In addition the examples provided some ideas for the statistical methods which may be used for analysing the position data.

4.2 Specification of Data

In contrast to the event data, the position data was not directly accessible from the smartfloor database. The position data can be viewed by the staff at Skovhuset, as animations in the user interface. For a particular room and time period, the user interface can present a video showing "dots" representing one or more people moving about a diagram of the apartment. Figure 17 shows a frame taken from such a video. The video includes the time and highlights when any events occur. At present, the only way to obtain the position data was in the form of these videos.
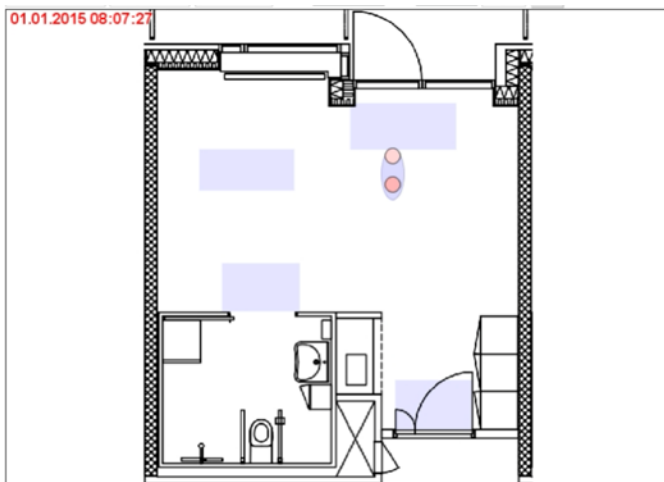

Figure 17 Example of a frame from an animation of a resident in their apartment presented in the smartfloor user interface

Therefore, the first step was to extract the position data from such a video. A video showing the movements within a single apartment over a 48 hour period from 08:00 on 01/01/2015 to 08:00 on 03/01/2015 was already available from a previous project (Capion, Gensby, Bloch, & Christiansen, 2015). The aim was therefore to extract the position of any dots appearing in the video along with the time. It was observed that the dots could not appear at any point in the room, but were limited to a grid of fixed possible positions. These are thought to correspond to the sensor-tiles that make up the smartfloor. Using image analysis techniques, each frame of the video was analysed so that

1. The tiles which are "active" in that frame (i.e. a dot is present in those tiles' position) are identified.
2. The time stamp of the frame is read off using OCR (optical character recognition).

This resulted in the raw dataset, which provided the time for each frame of the video, and the active tiles for each frame of the video.

4.3   Summary data from image analysis

There were 14134 frames in the video in which some tiles were active in the room. The video skips over periods where there is no activity in the room, so there were only a few frames which contained no active tiles, and these were discarded. Therefore, there were 14134 observations of time along with the active tiles at each time. The time between observations was less than or equal to two seconds for 90% of the observations. The remaining 10 % displayed a time difference that ranged up to 7 hours 42 minutes. The longest time differences occurred during the night time. In Figure 18 the time of each observation is shown on a timeline, and there are a number of long gaps in time between the observations. During these intervals there is no activity in the room (either the room is empty or the occupant is in bed).
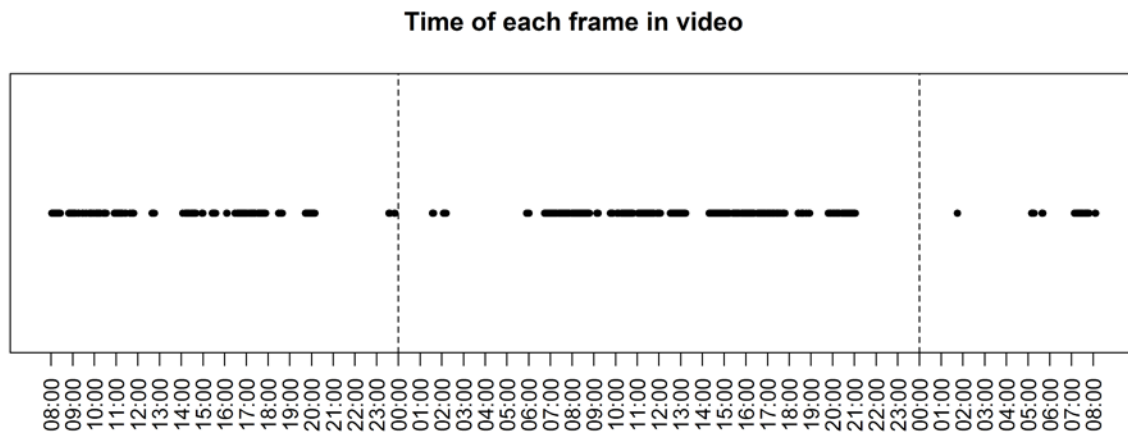

Figure 18 The time of each frame in the video. The vertical dashed lines indicate the start of a new day.

Next, the position variables are considered, which show which tiles are active for each frame. In Figure 19, the tiles are shown on a simplified diagram of the apartment with the floor tiles indicated by dashed lines. Each tile is labelled with the number of frames for which that tile is active (unlabelled tiles were not active in any of the frames). The magnitude of the counts is represented by coloured circles to form a heat map of the room. This illustrates where the greatest amount of activity took place in the room (green/yellow/red tiles) and where not much activity took place (purple/blue/empty tiles). There was a lot of activity around the entrance to the apartment as well as around the bathroom entrance. There was least activity in the "north" side of the apartment. The tiles which were never active in the center of the room are probably where the bed is located, and other furniture could be the reason for empty tiles around the edge of the room. Note that the plot indicates some active tiles located outside of the room. This is because the image analysis performed on each frame is not perfect, and will sometimes identify a "dot" which is not actually there, whilst other times it may fail to identify a dot which was actually present. Even so, the image processing was observed to be highly accurate, and the few active tiles it has identified that lie outside the room can be disregarded.
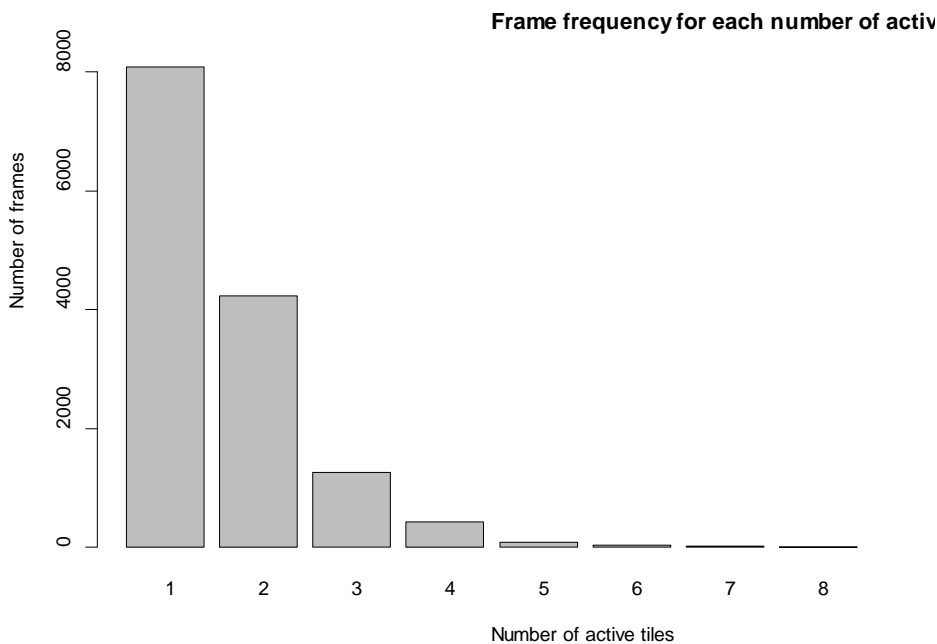
Figure 19 Heatmap of the activity in the room based on the frame frequency for each tile

Next, the number of tiles active for each observation is examined. Figure 20 shows that for 8000 frames only 1 tile is active, whilst for 4000 frames 2 tiles are active. There are rarely more than 3 active tiles in a frame. A single person is expected to activate 1 or 2 tiles when standing on the floor. If a greater number of tiles are active, maybe there are several people in the room at the same time. Figure 20 shows that for most of the time there are no more than 3 tiles active simultaneously and so it is probably reasonable to assume there is only one person in the room for the period covered by the video.

Figure 20 Number of frames for which different numbers of tiles were active

The data was processed further to create a format that was more convenient for analysis. Instead, of specifying which tiles are active at each time, it is better to have a single location coordinate for each time stamp, in the form of x and y coordinates. A simple coordinate system was used in which 1 unit was equal to 1 tile on the x and y axis (i.e. assuming the tiles are square). An observation with only 1 active tile is converted to the xy coordinate corresponding to the center of the tile. If several tiles are active, the xy coordinate of a tile center closest to the mean of the xy coordinates of the active tiles is used. In this way, each observation was converted to a point in three dimensional space of (x,y,time), ready for the next step of trajectory analysis.

4.4 Converting position data to trajectories and exploratory plots of trajectories

The first step in the trajectory analysis was to determine how to split the data into separate trajectories. Each trajectory should be a collection of position-time observations that belong to the same flow of movement or walk. It was decided to split the data into trajectories based on the time gaps between observations. Longer gaps occur between observations when there is no activity registered by the floor such as when the resident has left the room, is in the bathroom, is in bed or perhaps is sitting in a chair. The gaps are therefore natural endpoints for their walks around the room, and represent the end of one activity and the start of another. As previously discussed, most of the observations occur at time intervals of 1 to 2 seconds. It was decided to split the data into separate trajectories each time there was a time difference of 15 seconds between observations. This resulted in 392 separate trajectories.

In Figure 21, the distributions of various properties of the trajectories are shown. The number of observations within each trajectory varies greatly. There are many trajectories consisting of only 1 observation. These are caused by a floor tile being active at a moment in time, with no active floor tiles 15 seconds before or after

this moment. Perhaps if the resident is in bed and touches the floor briefly this could occur. Is may be that most of the trajectories of 1 observation are errors in either the smartfloor system, or the image analysis processing. The distance covered in each trajectory was calculated as the sum of the Euclidean distances between the positions coordinates of adjacent observations. The units of the distance are therefore in number of tiles. The duration of a trajectory is given by the time difference between the first and last observation in the trajectory. The average speed is then given by the ratio of the distance to the duration (so for speed to be defined, duration must be greater than 0, i.e. there must be more than one observation). There was a single trajectory with extreme values of 665 observations, 19 minutes duration and distance of 375 tiles. This trajectory is not included in Figure 21 as it would make the plots less detailed. As shown in Figure 21, all other trajectories contained less than 220 observations, lasted under 8 minutes and covered distances under 180 tiles.
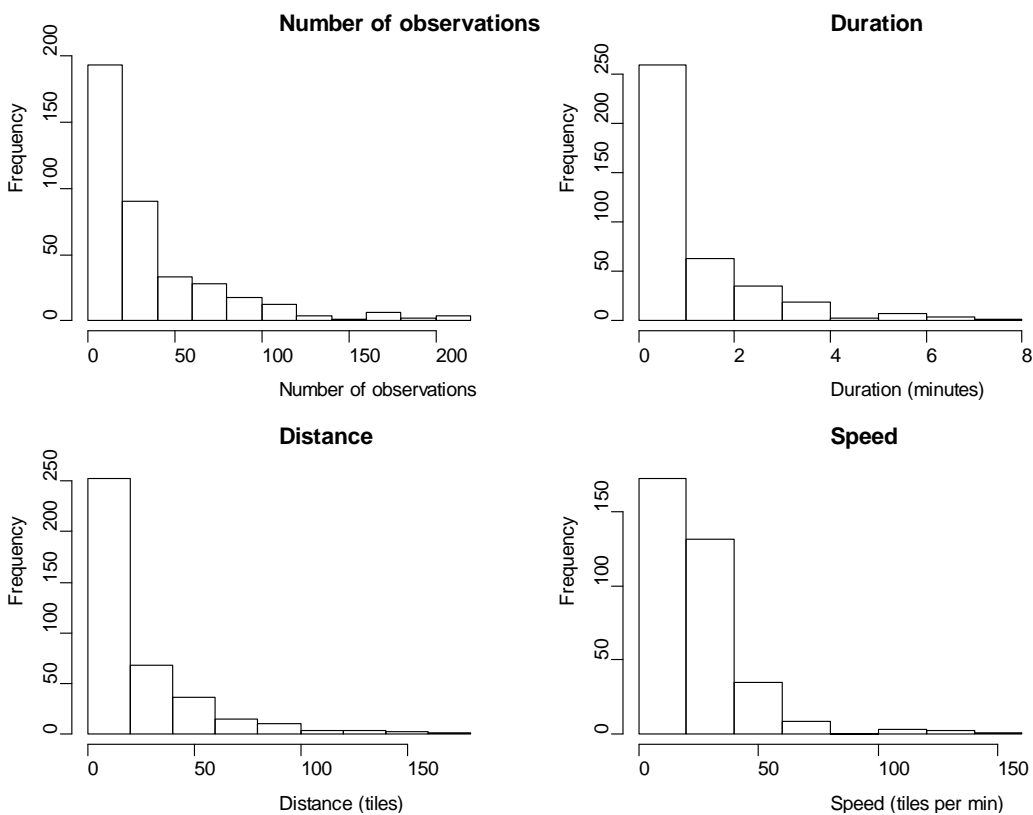

Figure 21 Some properties of the trajectories

The 37 trajectories consisting of a single observation are not of much interest as they are simply isolated position observations. Similarly, it was decided to discard the 80 trajectories of less than 5 observations.

The properties of the trajectories were examined over the course of the 48 hours. In Figure 22, the number of trajectories for each hour, and the total distance covered by the trajectories during each hour is plotted. This gives an impression of the general level of activity over the 48 hours. As expected, there is not much activity during night time hours. During the day there is a periodic pattern to the activity level so that periods of high activity alternate with periods of low activity.
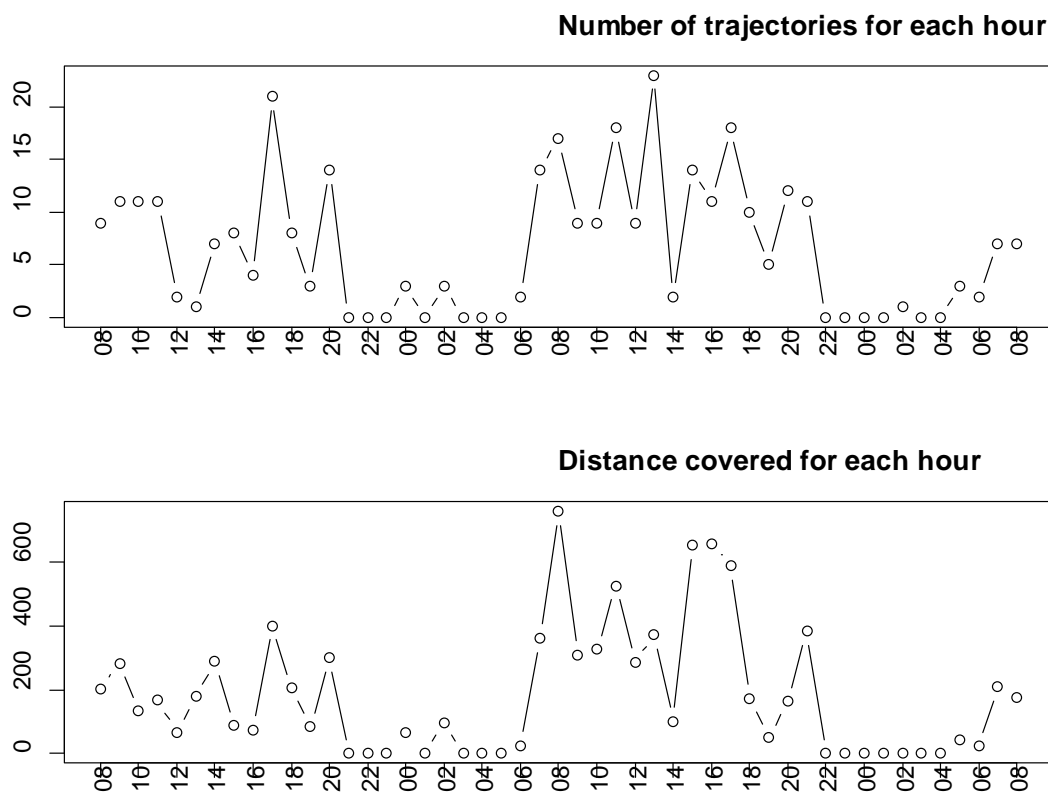
Figure 22 Number of trajectories for each hour and distance covered during each hour

Having examined some properties of the trajectories, the next step is to visualise individual trajectories. One method is a line plot of the position coordinates of the trajectory. In the data, the position coordinates are limited to being the center of the tiles, and so a lot of movement would be obscured by overplotting if the points are plotted directly as in Figure 23 (left). Therefore, the position coordinates are randomly "jittered" slightly (Figure 23, right), so that if a trajectory remains on the same tile for a period, this will be made visible in the plot. It should be remembered that the small perturbations in the plot are introduced for visualisation purposes. In addition, green and red dots are added to the plot to indicating the starting point and ending point respectively of the trajectory. Finally, a blue colour scale is used to indicate the speed of each "step" in the trajectory (ratio of change in distance to change in time between adjacent observations). Slow steps are shown in dark blue and faster steps are shown in lighter blue.
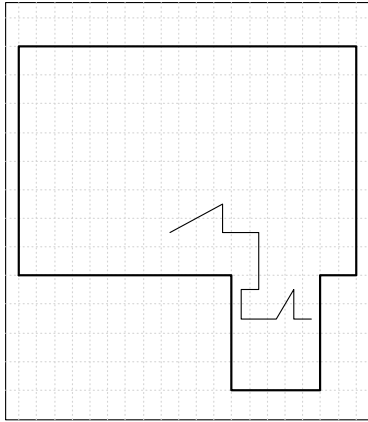
9, Day01 08:20:05   **9, Day01 08:20:05,**
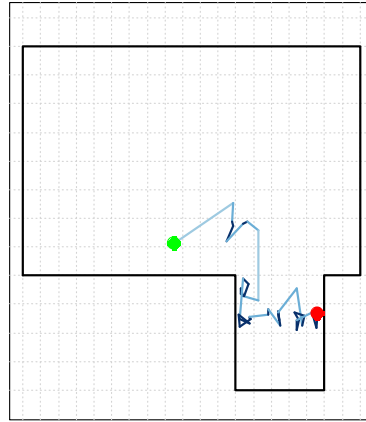
Figure 23 Plotting the same trajectory without jittering (left) and with jittering (right). The green and red dots indicate the start point and endpoint of the trajectory. The trajectory is coloured according to the speed between each observation (dark blue for slow, to lighter blue for fast). The plot title shows the trajectory ID number, day and start time.

Figure 24 presents 9 randomly selected trajectories plotted as described above. As well as occurring in different areas of the room, the trajectories also vary in length. Some of the trajectories display a lot of "lingering" (such as 189 and 222) where many steps stay on the same tile, whilst other trajectories (such as 27 and 231) show a more directed path.
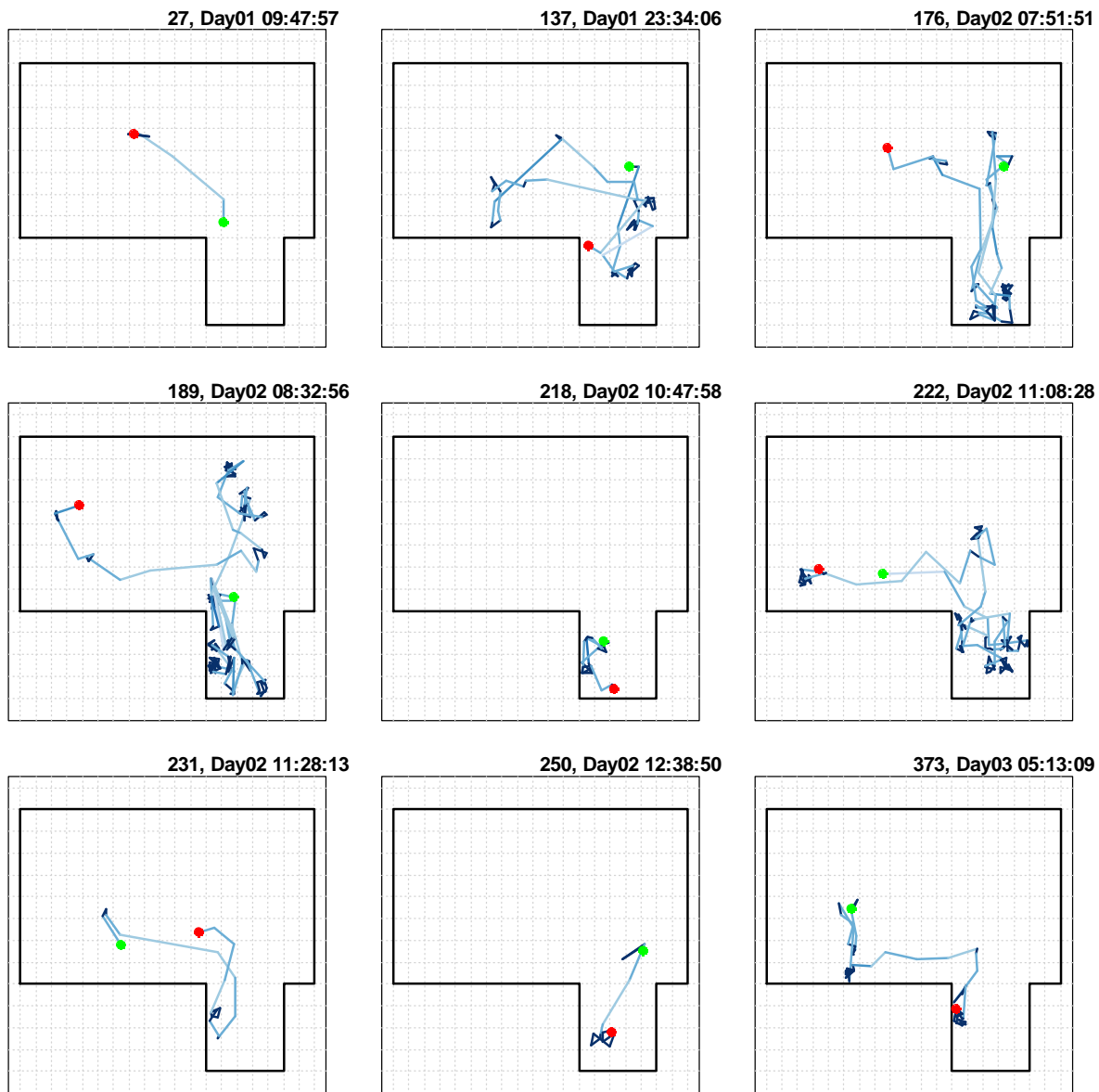
Figure 24 Plots of 9 randomly chosen trajectories

Another property of the trajectories which is of interest is the amount of change in direction that occurs during the trajectory. We refer to the line that connects one position obervation to the next position observation in a trajectory as a "step". Assuming that the resident always faces forwards in the direction of the step, then this direction is given by the angle between the step and the x axis. For convenience, we refer to the top of the room as dipicted in the figures, as facing North. For a step that remains on the same tile, we assume the same orientation as the most recent movement. A trajectory that consists of steps of many different orientations is one that often changes direction, whilst if all steps are of the same orientation, the trajectory is in a straight line. This can be represented graphically using radial histograms which show the relative frequency of steps for each angle interval. Radial histograms are presented in Figure 25 for the same nine trajectories plotted in Figure 24. By comparing the two figures it can be seen how the radial histograms summarise the direction property of each trajectory. For example, trajectory with ID 250 consists almost entirely of southwards movement.
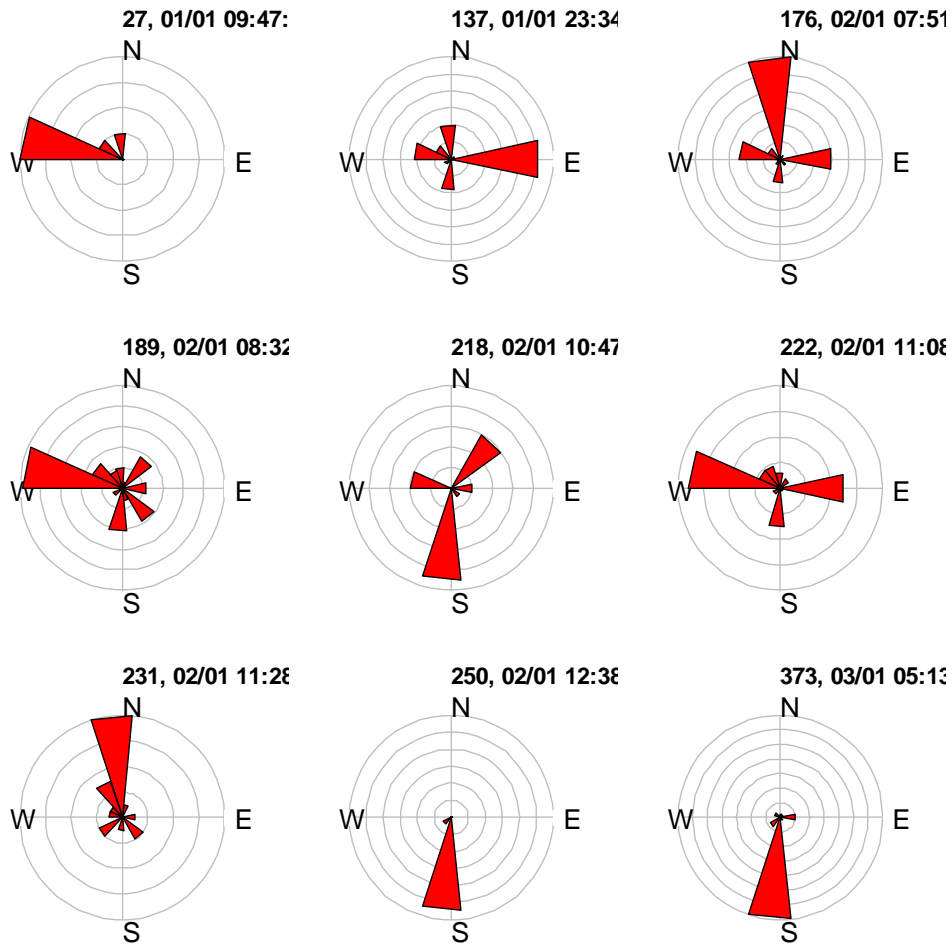
Figure 25 Radial histograms of step orientation for 9 trajectories. Within each plot, the amount by which each red slice extends indicates the relative frequency of steps taken in the directions indicated by that slice.

Another way to represent the trajectories is using a 3D plot as in Figure 26. This makes the time dimension more explicit and periods during which the resident is stationary appear as the vertical parts of the plotted line.
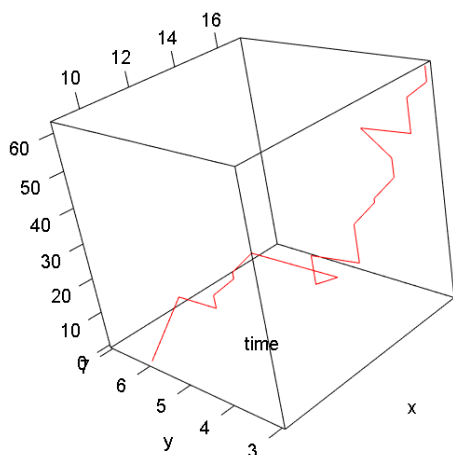

Figure 26 3D plot of trajectory 9

4.5   Analysis of trajectories

The aim of the analysis of the trajectory data was to investigate methods for comparing different trajectories to each other and that can determine to what degree trajectories are similar or different. The method used was adapted from Porikli, 2004, and involved the following steps:

- Each trajectory is standardised so that the observations occur uniformly every 2 seconds.
- The position observations are converted to a sequence of states, where each state is a region of the room.
- A Markov model is fitted to each trajectory sequence.
- To compare trajectories, the likelihood of the trajectories according to their own and each other's fitted Markov models is compared.

The aim of the first step was to ensure that the data for each trajectory was of the same format. However, it is not a necessary step for the method to be used. It should be noted that the data is recreated from video recordings and the time between observations is an artefact of the frame rate of the video. If in the future direct access to the position data is obtained then the time between observations is likely to be uniform. Standardising the data resulted in some observations being discarded, and some observations being repeated such that they occurred at 2 second intervals for the duration of each trajectory.

In order to convert the trajectories to a sequence of states, first the states must be defined. Each state is a region of the room. One option would be to define the states manually based on functions of different room areas such as, for example, the kitchen/entry area, the living area and the sleeping area. However, the number of such areas to define as well as their boundaries is somewhat subjective. We chose to define the states based on the position data in the trajectories. It is reasonable to assume that observations will cluster around the different functional areas and be sparser in between these areas, as is somewhat apparent in Figure 19. A method known as K-means clustering was applied to identify the centres of these clusters, and assign each observation to a cluster. In order to choose the number of clusters to use, k-means clustering is carried out for 1 to 15 clusters and the sum of squared errors (squared distances of each observation to their cluster's center) is examined (Figure 27). The greater the number of clusters, the smaller the sum of squared errors. However, beyond 6 clusters, the sum of squares does not get much smaller when additional clusters are used. Therefore, a reasonable number of clusters to use is 6.
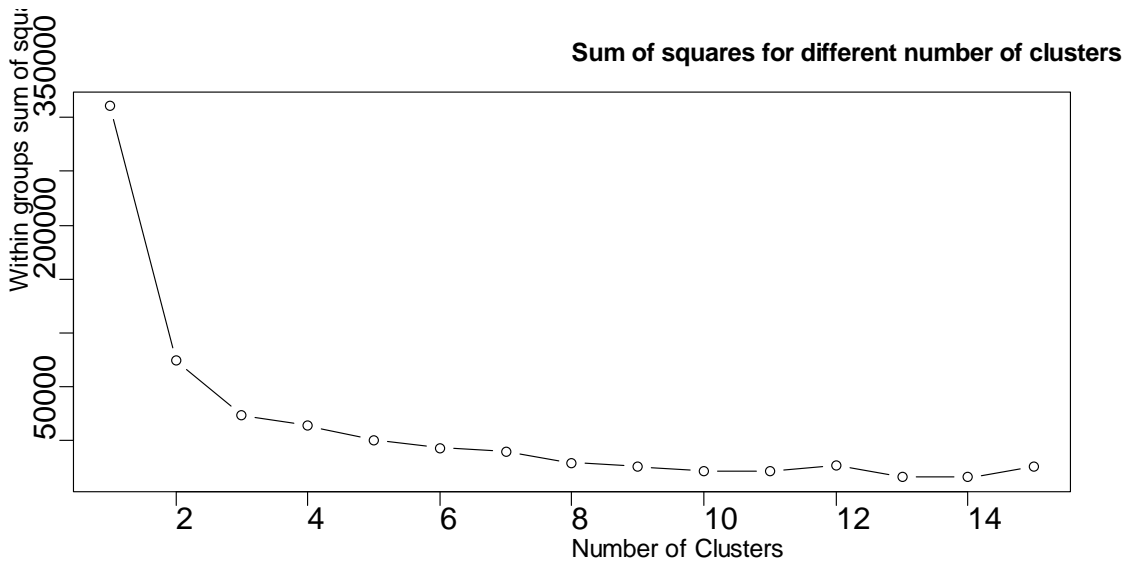
Figure 27 Within groups sum of squares for K-means clustering using different number of clusters

After applying k-means clustering with 6 clusters to the data, the room is divided into 6 regions as indicated in Figure 28. This means that each observation is assigned to the state number to which its position belongs.
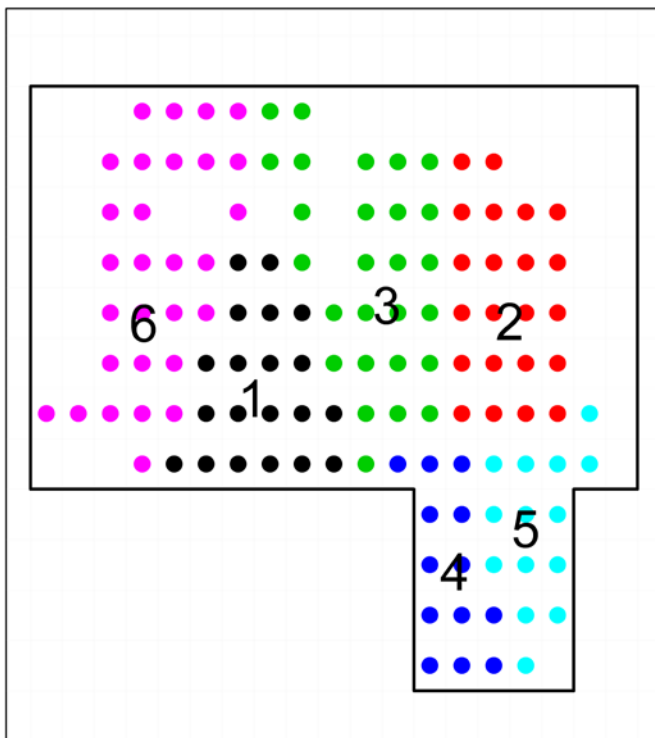

Figure 28 The 6 clusters that resulted from applying k-means clustering

The trajectories can then be represented as a sequence of states, rather than the position-time coordinates used previously. For example, trajectory 27 shown in Figure 24 (top left) is represented by the state

sequence "4, 3, 1, 1, 1, 1". By converting the trajectories to state sequences, their complexity is greatly reduced.

The next step was to represent state sequence of each trajectory by a first order Markov model. In this model, the probability that the next observation in the sequence will take on each state depends only on the current observations state. If we assume that all the trajectories are generated by the same process, then the parameters of the first order Markov model that is fitted to the entire data is displayed in Figure 29. According to this model, given that a sequence is in state 3, the probability that the next observation will also be in state 3 is 0.77, whilst the probability that the next observation will have state 1 is 0.08.
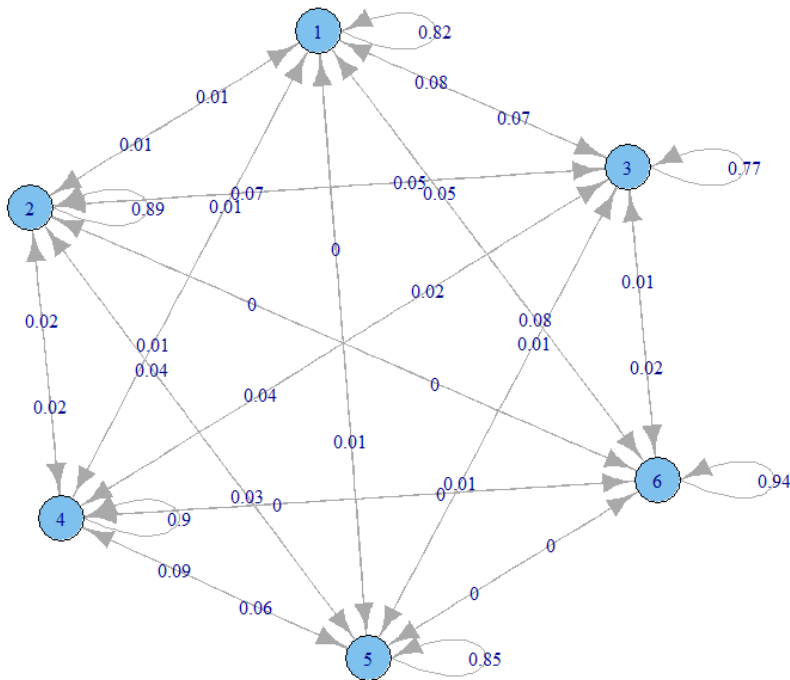

Figure 29 The parameters of the Markov model fitted to all of the data

A separate Markov model is fitted to the state sequence of each trajectory. This is done by simply counting the number of transitions from state to state, within the sequence. However, if there are no transitions from one particular state to another particular state, the probability estimate would be 0. To avoid probability estimates of 0, additive smoothing is used to estimate the probabilities. Formally, the probability of observing state *j,* after just having observed state *i*, is estimated using the formula

$$p_{ij} = \frac{n_{ij} + \lambda}{\sum_i n_{ij} + 6\lambda}$$

Where n_{ij} is the count of transitions from i to j in the sequence, and lambda is the additive smoothing constant (a value of 0.1 was used).

Having estimated the Markov parameters for each trajectory, the distance metric described in Porikli (2004), can be used. This distance metric is defined as follows. Let T_i denote trajectory i, with corresponding estimated Markov parameters θ_i . Then the distance between two trajectories, T_1 and T_2 is given by

$$dist(T_1, T_2) = |\, l(T_1|\theta_1) + l(T_2|\theta_2) - l(T_1|\theta_2) - l(T_2|\theta_1)\,|$$

Where $l(.)$ Denotes the log likelihood function. In other words, this distance metric is the log likelihood ratio of the likelihood of the trajectory sequences given their own Markov paramenters, to the likelihood of the sequences given eachothers' Markov parameters. If the two trajectories are identical then all the terms in the

expression will be the same and the distance will be 0. If the two trajectories are very different, then their Markov parameters will differe greatly, leading to a greater difference in likelihoods and a greater distance value.

For each trajectory, the distance was calculated between that trajectory and every other trajectory, and the trajectory with the smallest distance was identified. In this way, each trajectory could be paired with the most similar trajectory in the data set. Examples of the given trajectories and their "nearest neighbour" trajectories are shown in Figure 30, in order of distance to nearest neighbour. That is, the first plots are of trajectories for which the nearest neighbour was very close, whilst in later plots the nearest neighbour was farther away from the given trajectory. From examining Figure 30, and the same plots made for the other trajectories, it appears that the distance metric performs well at evaluating the overall shape of the trajectories. The method can evaluate the similarity of trajectories based on overall shape and location, and will also tend to match trajectories of similar length. The method appears to be fairly robust, and wil ignore small differences between the trajectories. As long as two trajectories have the same general shape (e.g. left to right movement) and are located in the same area of the room, then the distance metric will evaluate them is close together.

In Figure 31 (left) the distribution is shown of the distance of every trajectory to its nearest neighbour trajectory according to the distance metric. Trajectories which are relatively far away from a nearest neighbour could be regarded as rather isolated occurrences. For example, the red line in the figure marks the distance to nearest neighbour value which is only exceeded in 1% of the trajectories and these could be seen as unusual or out of character for the resident. However, it should be noted that the dataset only covers less 48 hours of the residents movements. Data for a much longer period would be need to reliably identify usual and unusual trajectories. Figure 31 shows that all of the trajectories were within a distance of 50 from their nearest neighbour, except for one trajectory which was a distance of nearly 70 from its nearest neighbour, and this was the trajectory mentioned previously which was much longer than all of the others.

The three most unusual trajectories according to the distance metric are shown in Figure 32. They all appear to be very long trajectories. In Figure 31 (right), distance to nearest neighbour is plotted against trajectory length (number of observations in the trajectory). The plot indicates that there is a possitive correlation between trajectory length and distance to nearest neighbour. This may be explained by the observation that the longer and more complex a trajectory is, the less likely a similar trajectory will occur.
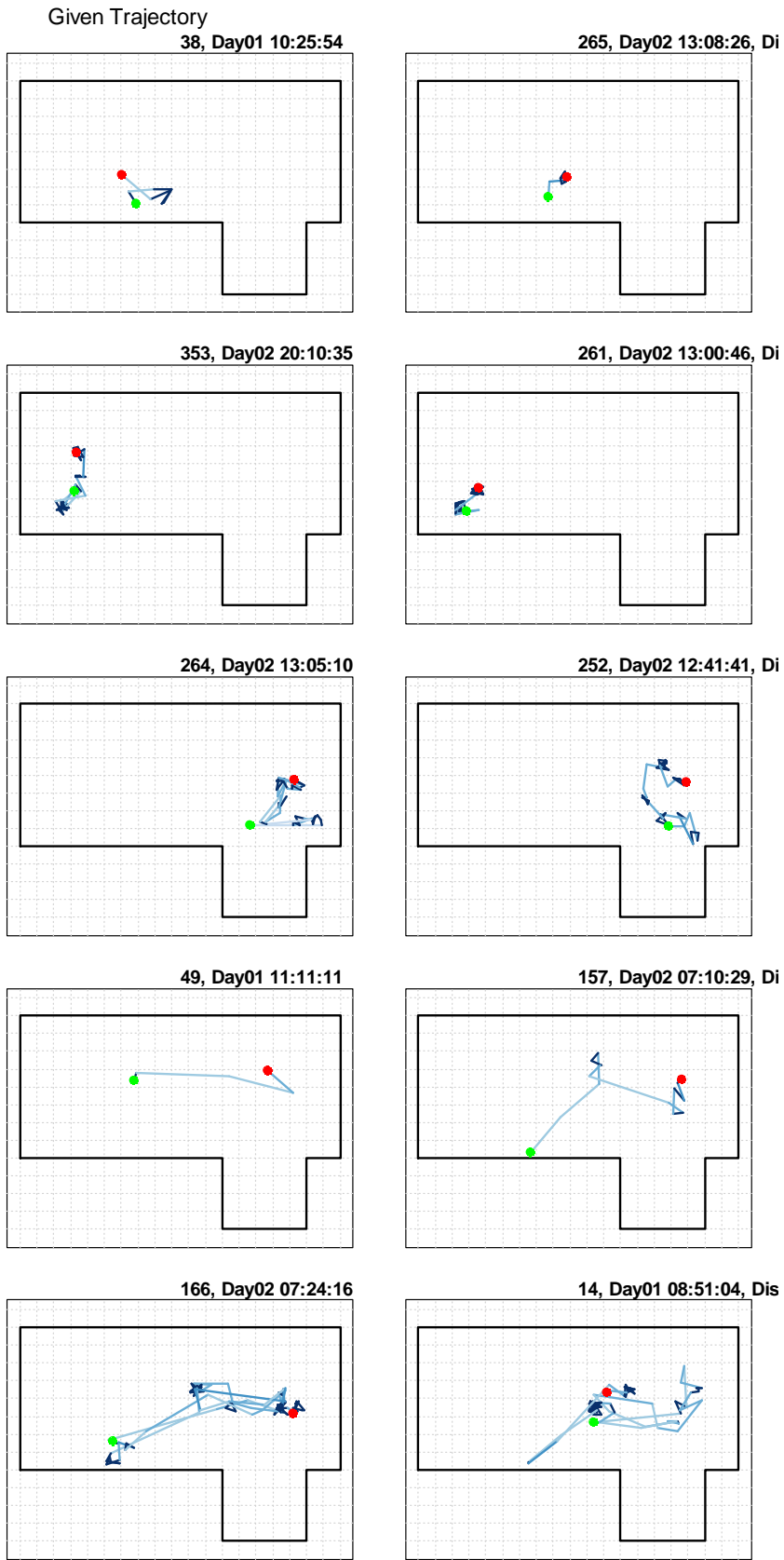
Given Trajectory

Figure 30 For each of the five given trajectories on the left, the most similar trajectory according to the distance metric is plotted on the right. The distance between each left and right trajectory is included in the title of the right trajectory
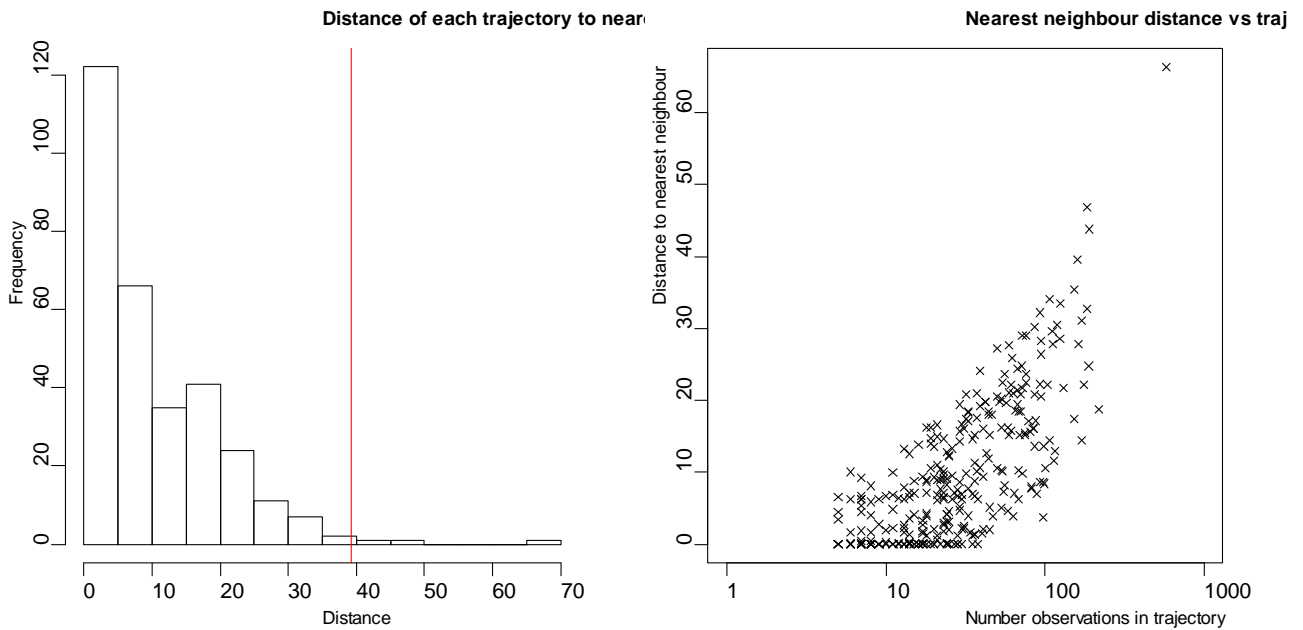
Figure 31 Histogram of the distance of each trajectory to its nearest neighbour (left). The red line indicates the 99% quantile of the distances. Scatter plot of the distances against number of observations in the trajectories.
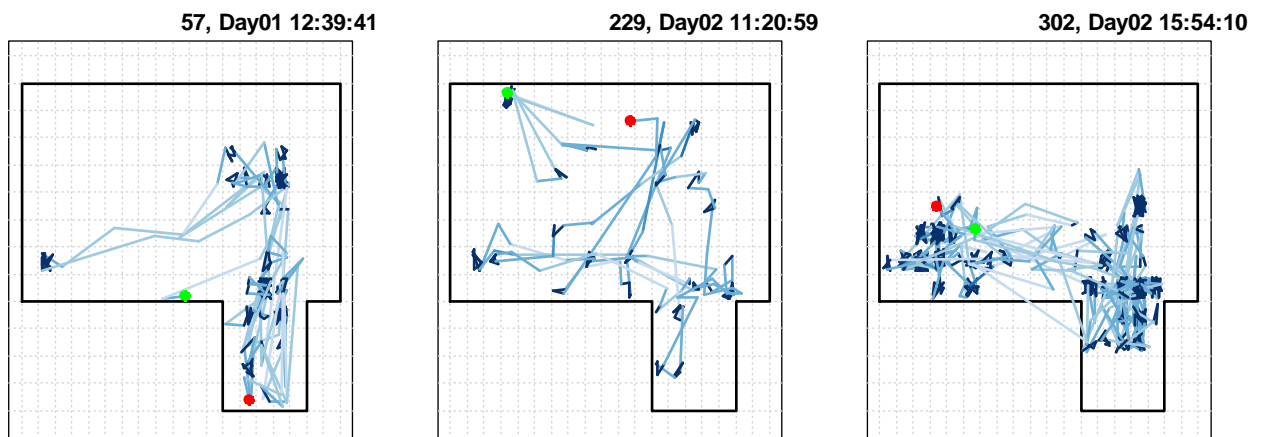

Figure 32 The three trajectories which had the greatest distance to a nearest neighbour

5. Conclusion

This project has demonstrated new ways of making use of data from the smartfloor at Skovhuset care home. The application of control charts to the event data illustrated how it may be possible to foresee UTI incidents based on the number of ENTERWC and EXITBED events during the night time, however a definite link was not confirmed. Even so, it was shown how control charts can provide useful insight into the behaviour of the residents over time, and how they can identify changes in behaviour. This information would help the staff provide better care to the residents. The information would be useful for diagnosis as well as enabling resources to be focused on cases where unusual activity is observed. A limitation of event data control chart approach is that the event data is not always guaranteed to be generated by the resident, but can often be caused by visitors or staff. This could be a reason why the night time event counts may provide a more accurate reflection of the resident's behaviour.

There are many opportunities for further investigation into applying control charts to the event data. In this project, only three examples of residents treated for UTI were investigated. A more thorough investigation could be carried out if clinical data for more residents was made available. Another possibility would be to investigate whether falls can be predicted. Ideally, this would require information on falls to be recorded by Skovhuset as the fall event in the data may contain many false alarms. Further work could also be carried out to investigate using multivariate control charts, which could be better at providing a combined view of behaviour using all the event variables.

The analysis of the position data also demonstrated that properties of the trajectories traced by the residents on the floor can be examined over time for changes in behaviour. A method was found for objectively comparing trajectories and measuring how unusual they are. The position data was obtained indirectly using image analysis on the smartfloor video animations, and so is an approximation of the position data contained within the smartfloor database. Even so, the methods were shown to be effective and could easily be applied to the "real" data at a later date. The distance metric used in this project was reasonably straight forward and fast to implement. Further work could investigate more sophisticated trajectory distance measures such as the "Fréchet distance" (see for example Khoshaein, 2014). If future work is carried out into the smartfloor position data, then it would be nessary to obtain direct access to the data so that position data for several residents over a long period can be analysed. This could reveal greater insight into what properties of the position data can be used for improving the care provided to the residents. It is also likely that data obtained directly from the smartfloor contains additional information such as pressure intensities on the floor tiles and this would provide another line of investigation.

6. References

Capion, J., Gensby, E., Bloch, S., & Christiansen, A. (2015). *Sensorgulv i nyt plejecenter.* Lyngby: DTU.

Gellert, A., & Vintan, L. (2006). Person movement prediction using hidden Markov models. *Studies in Informatics and control* , 15.1, 17.

Khoshaein, V. (2014). *Trajectory Clustering using a Variation of Fréchet Distance (Thesis).* University of Ottawa.

Majecka, B. (2009). *Statistical models of pedestrian behaviour in the Forum.* University of Edinburgh.

Porikli, F. (2004). Trajectory distance metric using hidden markov model based representation. *IEEE European Conference on Computer Vision, PETS Workshop*, Vol. 3.

Stauffer, C., & Grimson, W. (2000). Learning Patterns of Activity Using Real-Time Tracking. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, 22, 747–757.

Suzuki, N., Hirasawa, K., Tanaka, K., Kobayashi, Y., Sato, Y., & Fujino, Y. (2007). Learning motion patterns and anomaly detection by Human trajectory analysis. *IEEE International Conference on Systems, Man and Cybernetics*, 498-503.

Woodall, W. H. (2006). The use of control charts in health-care and public-health surveillance. *Journal of Quality Technology*, 38.2: 89-104.

Xie, M., & Thong, G. (2006). Some Statistical Models for the Monitoring of High-Quality Processes. *Springer Handbook of Engineering Statistics*, 281-290.
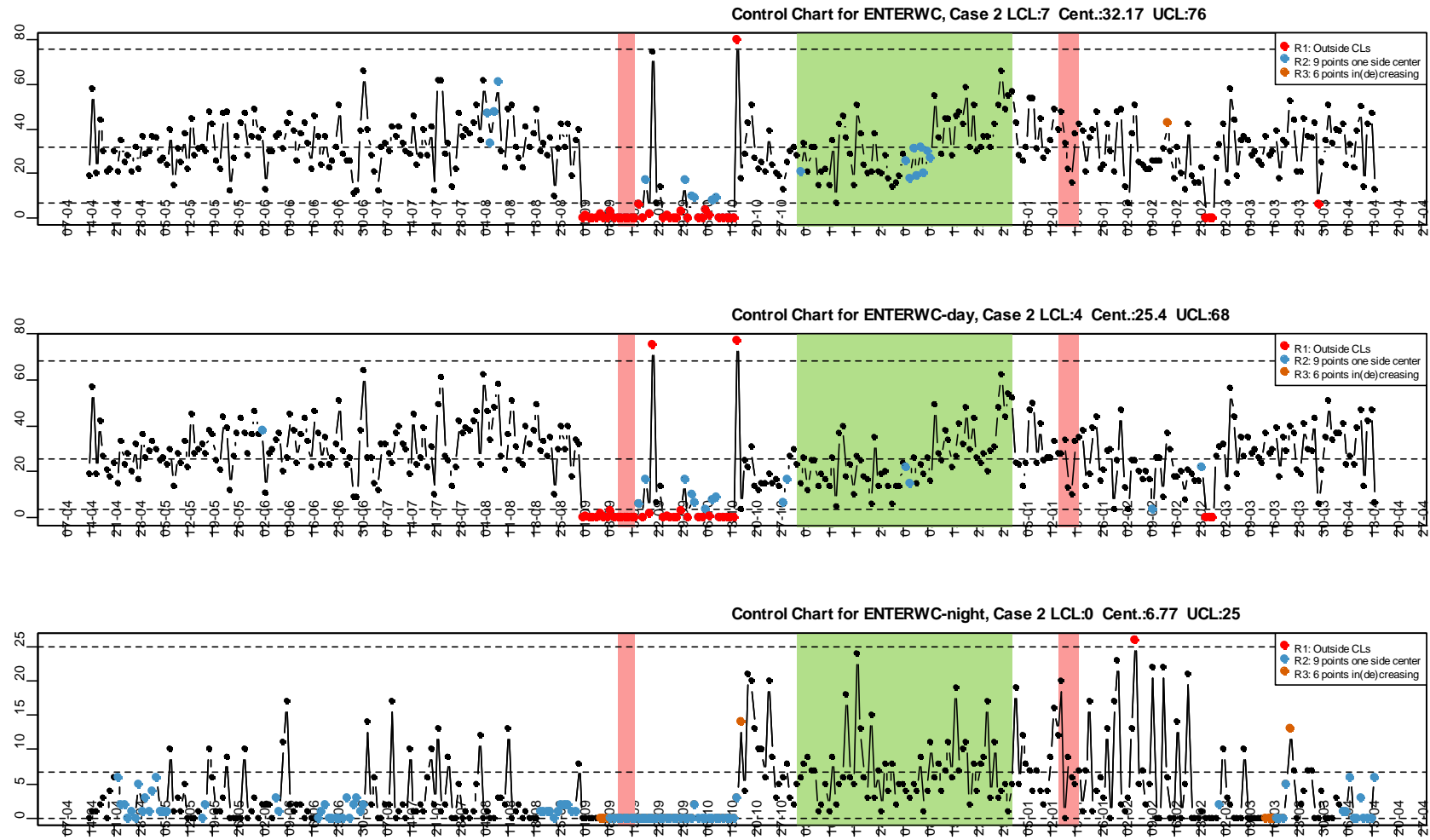
7. Appendix

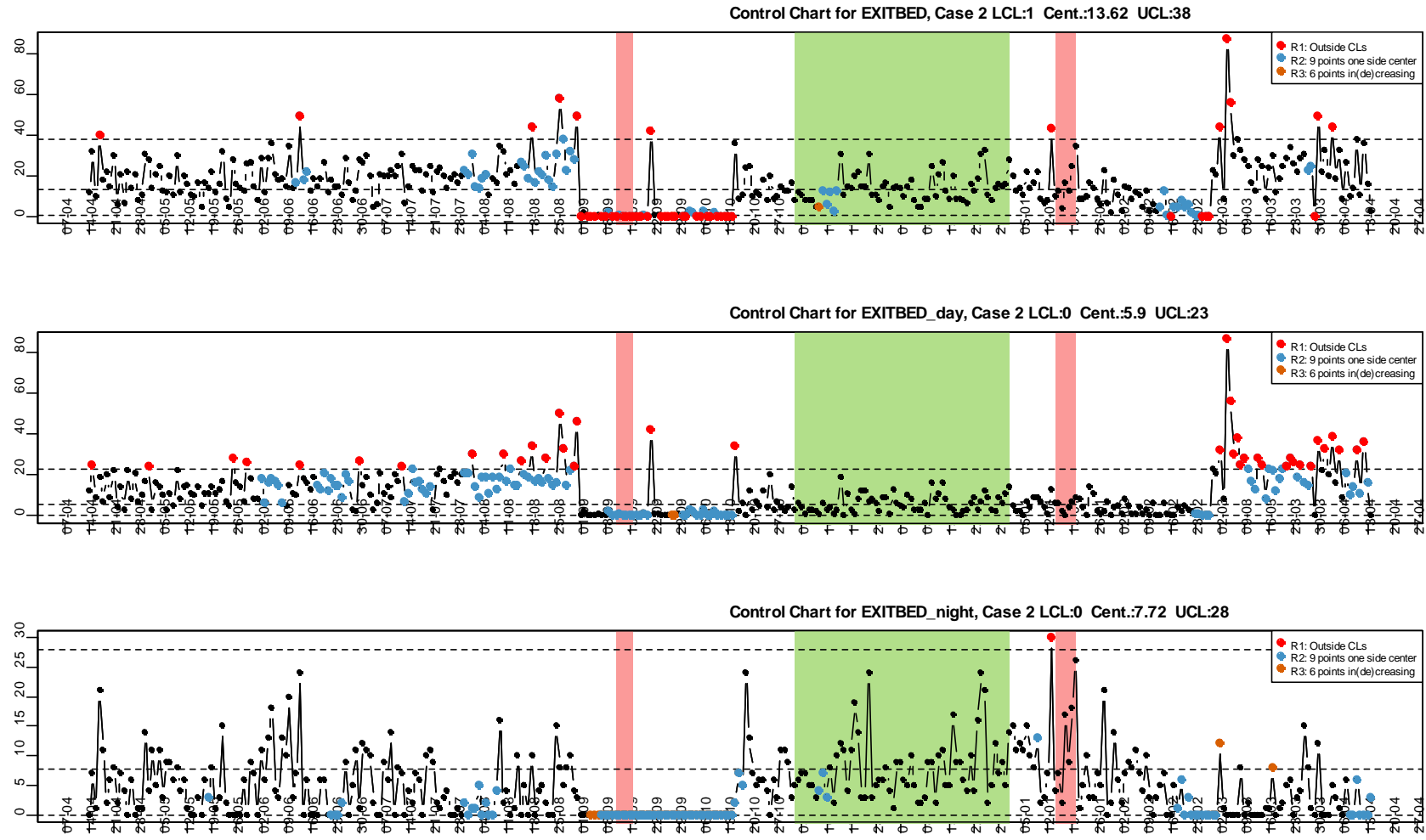**Figure 33 Control charts of ENTERWC, ENTERWC-day and ENTERWC-night for Case 2**

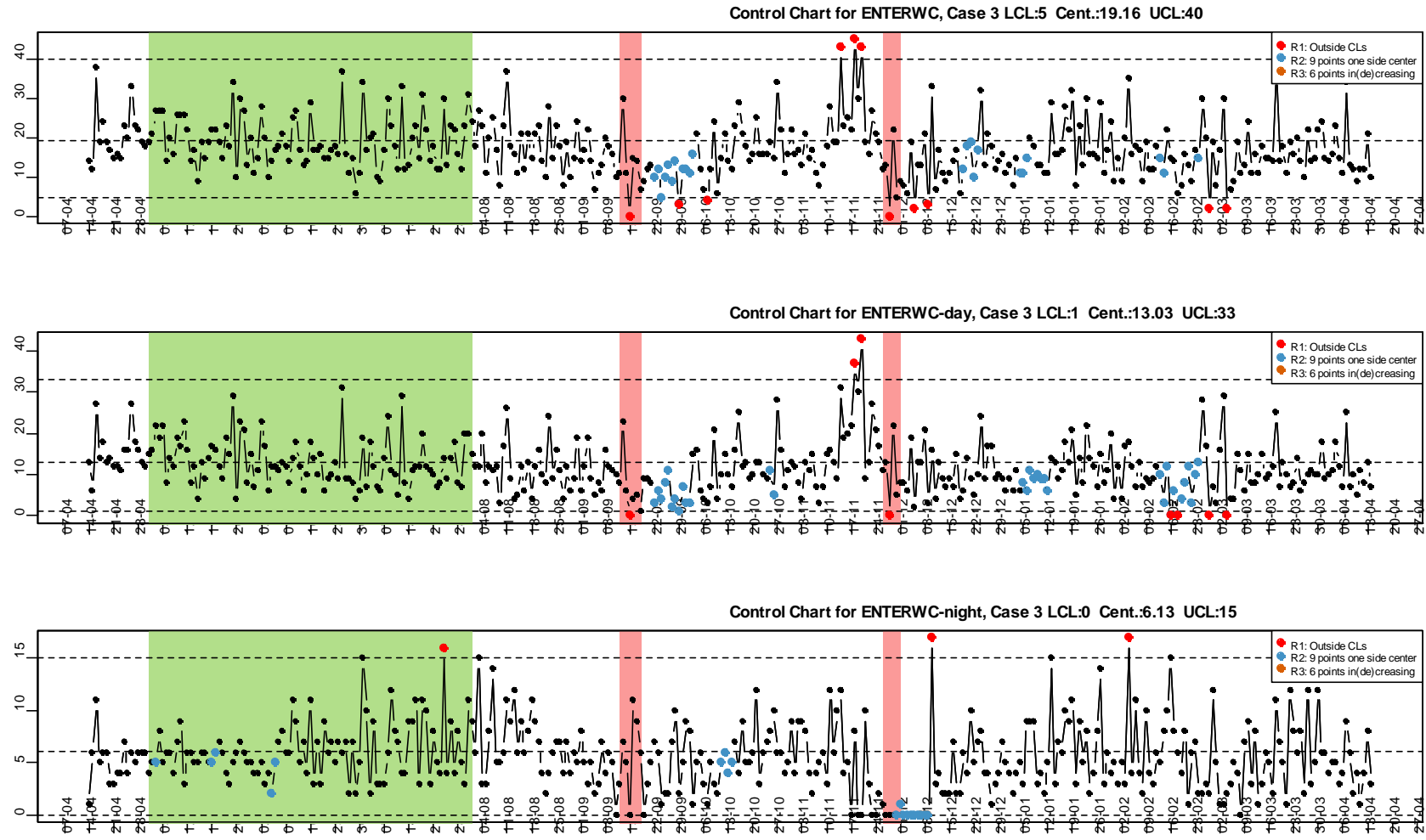**Figure 34 Control charts for EXITBED, EXITBED-day and EXITBED-night for case 2**

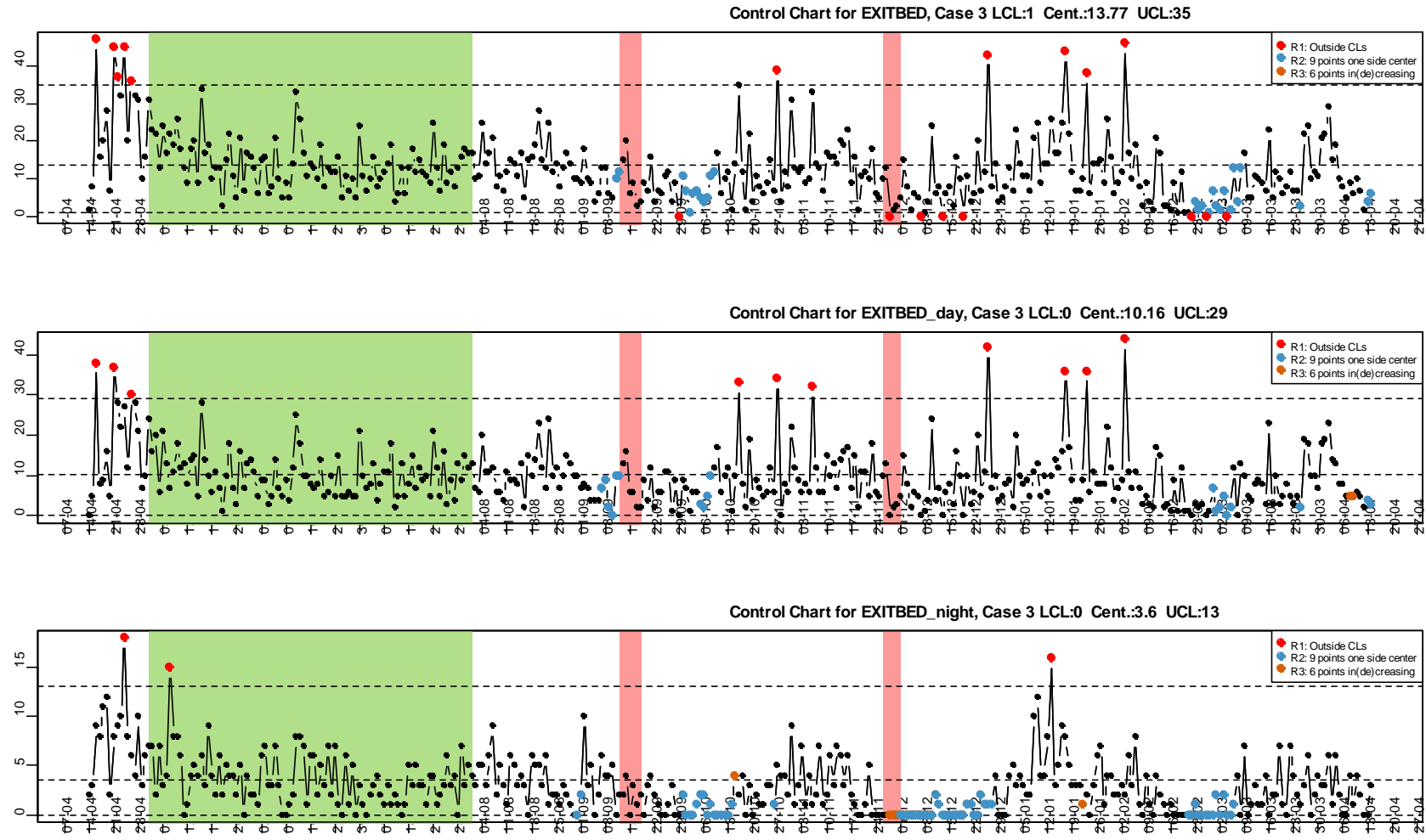**Figure 35 Control charts of ENTERWC, ENTERWC-day and ENTERWC-night for Case 3**

Figure 36 Control charts for EXITBED, EXITBED-day and EXITBED-night for case 3