Inferring Household Occupancy Patterns from Unlabelled Sensor Data

Wilhelm Kleiminger*[1], Christian Beckel†[1], Anind Dey‡[2] and Silvia Santini§[3]

[1]Institute for Pervasive Computing, ETH Zurich, Switzerland
[2]HCI Institute, Carnegie Mellon University, USA
[3]WSN Lab, TU Darmstadt, Germany

September, 2013

Abstract

This technical report describes the *homeset* algorithm, a simple yet effective approach to estimate home occupancy schedules from unlabelled sensor data. The algorithm relies on Wi-Fi scan data to determine when residents are at home and when not. We validate our approach using a data set from the Nokia Lausanne Data Collection Campaign that contains mobile phone traces of 38 participants collected over more than one year. Since the data is unlabelled, we indirectly validate our results leveraging the information hidden in anonymised GPS traces collected by the mobile phones of home occupants. We further show that the homeset algorithm is able to autonomously determine the reliability of the computed schedules. Finally, we show how these schedules can be used to predict the future occupancy behaviour of mobile phone owners.

1 Introduction

A number of studies have shown how behavioural patterns of both groups and individuals can be discovered by analysing data collected using off-the-shelf mobile devices [3]. For instance, mobile phones have often been used to gather mobility traces of individuals [3, 4]. The analysis of these traces enables, e.g. identification of places of interest in the daily lives of individuals [4] or even the prediction of places that will most likely be visited by the mobile phone holders [7].

The use of mobile phones for the collection of mobility traces thus makes it possible to explore, model and predict human behaviour. Retrieving mobility traces at a fine temporal and spatial scale, however, may consume a significant amount of resources. For instance, the continuous operation of GPS is known to shorten battery lifetime of mobile phones significantly [2]. In practical settings, the use of GPS is thus typically "rationed" and combined with other technologies, in particular cell- or Wi-Fi-based localisation [6]. This, however, also requires reliance on third-party services and might thus raise privacy issues.

To reduce the impact of these issues, collecting data at a much coarser scale might still be sufficient to support a large set of applications and at the same time preserve mobile phone resources and protect users' privacy. Such scenarios include applications that rely on knowledge about when households' occupants are likely to return home, like home automation applications (e.g. automatic heating control), location-based reminders or notification services to ensure the presence of children at home.

*wilhelmk@inf.ethz.ch
†beckel@inf.ethz.ch
‡anind@cs.cmu.edu
§santinis@wsn.tu-darmstadt.de

In this report, we focus on this specific class of applications and present a novel approach for discovering the daily occupancy patterns of private households using mobile phones. Our approach allows mobile phones to autonomously estimate the occupancy schedules of their owners. To this end, we leverage Wi-Fi scan traces, i.e. the information that mobile phones gather about visible Wi-Fi access points. The approach requires scans to be performed only at a coarse temporal scale (e.g. every 15 minutes). By performing a time-based clustering of these traces, our *homeset* algorithm can accurately reconstruct the probabilistic occupancy schedule – i.e. the probability of her mobile phone being at home at any given time and day – of each household's occupant. These probabilistic schedules can then be used to estimate future occupancy schedules and thus support the class of applications mentioned above.

We evaluate our approach using Wi-Fi scan traces gathered in the context of the Lausanne data collection campaign (LDCC). This campaign was launched in 2009 by the Nokia Research Center in Lausanne, Switzerland and the collected data was recently released in the context of the Nokia Mobile Data Challenge (MDC) [6]. The released data set contains more than one year worth of traces of Wi-Fi scans, GPS coordinates, accelerometer readings and several other sensors, as well as demographic information for 38 of the mobile phone users that participated in the data collection campaign. The participants are identified by the identifier used in the LDCC data set (i.e. participant "007").

As the data set does not contain information about places of interest of the LDCC participants (e.g. we have no information about where the "home" of the participants is), we evaluate our findings by developing a heuristic approach that leverages the GPS traces in the data set. In particular, we first select those GPS coordinates that have been truncated in order to obfuscate the actual position of the mobile phone. We then apply a temporal clustering procedure on these coordinates to observe in which time frames specific coordinates have been obfuscated. Finally, we infer the type of place (e.g. home or work) corresponding to a specific set of coordinates depending on when they have been truncated (e.g. coordinates corresponding to the home are typically obfuscated during the night). This approach allows us to gather the ground truth information necessary to empirically evaluate the reliability of our homeset algorithm – without any need to retrieve the actual location of the participants' homes.

After describing our method to derive probabilistic occupancy schedules and how we evaluated its performance, the report describes how the obtained schedules can be used. To this end, we first show how future occupancy schedules can be predicted relying on historical data. Further, we present preliminary results showing that it is possible to use selected features of the occupancy schedules to automatically recognise specific classes of individuals (e.g. full-time workers vs students).

Before presenting the homeset algorithm and discussing its performance in Section 3, we summarise related work in Section 2. We then present two examples on how the derived occupancy schedules can be used in Section 4. Section 5 concludes the report and presents our outlook for future work.

2 Related work

The idea of using mobile phones to discover human mobility patterns has been explored extensively in the last few years [3]. Several authors have focused on identifying places of interest (e.g. work-place, home) and on predicting transitions between such places [1, 4]. Our work is related to these approaches since we aim to identify – although not locate – the home of a mobile phone user in order to build a probabilistic schedule which estimates the probability that she will return home at a future time interval.

In [7], Scellato at al. address the problem of estimating the arrival time of a user at specific location as well as "*the interval of time spent in that location*". To this end, fine-grained location traces need to be collected and evaluated centrally. Similarly, the PlaceSense algorithm by Kim et al. [4] uses a sampling rate of 0.1 Hz to find semantic places, which resulted in short battery life-times that would not be suitable for a long term deployment. Furthermore, since PlaceSense

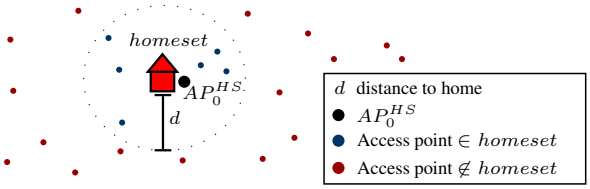


Figure 1: The homeset is the set of access points in the vicinity of the home access point AP_0^{HS}.

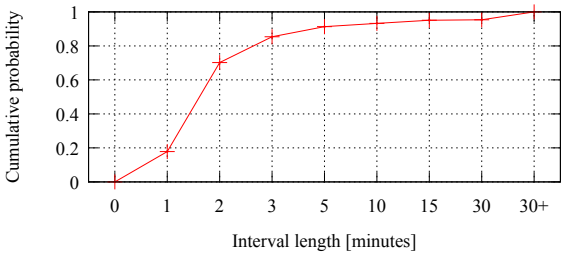


Figure 2: Cumulative probabilities for intervals from 1 minute to infinity across all participants.

requires a stable scan window (i.e. a consistent set of beacons) to detect entrance to a place, it struggles on data sets with relatively low sampling rates such as the MDC data set used in this report. In contrast, our homeset algorithm requires collecting only coarse-grained traces of Wi-Fi scans and can operate locally on the user's phone.

Other authors have also focussed on the problem of predicting the occupancy of private households in order to support home automation applications. Existing approaches to discover occupancy schedules mainly rely on the availability of data from a GPS logger to compute distance from home or ad-hoc sensors (e.g. passive infrared sensors) installed in the home [5, 8]. While the installation of ad-hoc sensors poses an additional burden in terms of costs and maintenance effort, the continuous operation of a GPS module is typically avoided due to energy constraints [6]. Thus GPS data is often replaced by or combined with information gathered through Wi-Fi- or GSM-based localisation services [4, 6]. Figure 1 shows a comparison of GPS based presence detection with our homeset algorithm. Related work [5] has put a user as *home* if she was in a 100 m radius of her home. We therefore argue that being within the coverage area of a Wi-Fi network is sufficient to detect occupancy at a much lower energy cost.

3 The Homeset Algorithm

The goal of the homeset algorithm is to compute the *probabilistic occupancy schedule* of the residence of a mobile phone owner. A schedule is represented as a matrix P with 7 columns, one for each day of the week and N_s rows. N_s is the number of temporal *slots* within a day. N_s can be set to an arbitrary value, depending on the desired time granularity of the schedules. Figure 2 shows that in the MDC data set, the interval between consecutive Wi-Fi scans is less than 15 minutes in 95% of the cases. In the context of this work we thus consider slots of 15 minutes, such that $N_s = 24 \times 60/15 = 96$.

As an example, figure 3 shows the probabilistic occupancy schedule derived for participant 007 of the MDC data set[1]. This schedule reveals that participant 007 is usually away from home between $8am$ and $7pm$ during weekdays, while her behaviour is far less regular on the weekends.

To compute the probabilistic occupancy schedules, the homeset algorithm relies on logs of Wi-Fi scans only. Each time a mobile phone detects the presence of a Wi-Fi access point it stores several pieces of information. Among these, the homeset algorithm only uses the timestamp of

[1]Figures 9 to 13 in the appendix show the schedules for all 38 participants included in this report.

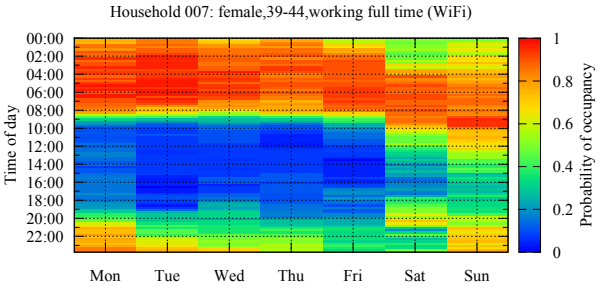


Figure 3: Probabilistic Wi-Fi occupancy schedule for participant 007. The participant is most likely not to be at home on weekdays between 8am and 7pm.

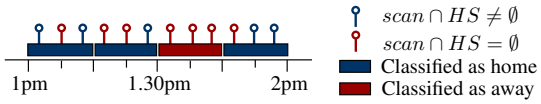


Figure 4: Interval classification based on multiple scans and homeset.

the scan and the MAC addresses of the visible access points. A single Wi-Fi scan is thus a tuple $< ts, AP_0, AP_1, \ldots, AP_{m-1} >$ where m is the total number of access points seen in a particular scan and AP_i is the MAC address of and thus uniquely identifies, a specific access point.

The homeset algorithm uses these scans to identify a set of access points that are located within, or in the immediate proximity of, the household of a mobile phone owner. We call this set the *homeset* (HS) and assume it contains n access points, so that $HS = \{AP_0^{HS}, AP_1^{HS}, ..., AP_{n-1}^{HS}\}$. We will for now assume $n > 1$ and discuss the initialisation of the HS below.

Figure 4 shows occupancy classification with the homeset algorithm. Given a Wi-Fi scan $< ts, AP_0, AP_1, \ldots, AP_{m-1} >$ the homeset algorithm tests whether $\{AP_0, AP_1, AP_2, ..., AP_{m-1}\} \cap HS \neq \emptyset$. If this statement returns true, the algorithm assumes the household to be occupied in the slot i of day j identified by the timestamp of the scan. This observation is then stored in a *occupancy frequency matrix* O of dimensions $N_s \times 7$, i.e. the element o_{ij} of matrix O is incremented by 1. Concurrently, a *total observations matrix* T is maintained and used to store the total number of times a Wi-Fi scan has been registered in a particular slot i of a day j. Note that both o_{ij} and t_{ij} are incremented at most once in each time slot.

The elements of the probabilistic occupancy schedule matrix P are then computed as:

$$p_{ij} = \begin{cases} o_{ij}/t_{ij} & \text{if } t_{ij} > 0 \\ 0.5 & \text{otherwise} \end{cases}$$

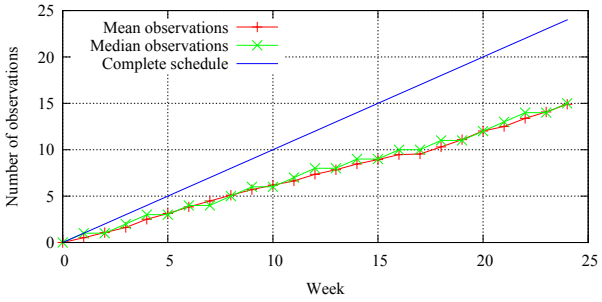
As indicated in this equation, as long as no scans are available, the homeset algorithm assumes there is an equal chance for a participant to be at home or away ($p_{ij} = 0.5$).

By repeating this procedure over a few weeks, the desired probabilistic schedules can be computed. We will discuss below how to automatically determine when a schedule can be considered *mature*.

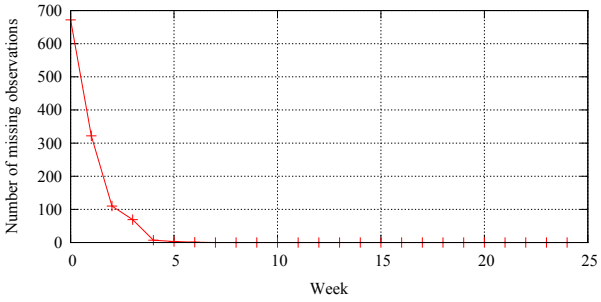
3.1 Initialisation of the homeset

In order to initialise the homeset in practical settings, one could require the user to manually enter the MAC address of the household's private access point, if one exists or to actively scan for nearby access points while at home.

To eliminate this manual effort in initializing the homeset from the available MDC data, we computed the the empirical probability ω_x of seeing access point x at least once between $3am$ and $4am$ on any particular night. This procedure relies on the assumption that people spend most of their nights at home. The access point with the highest value for ω_x is set to be AP_0^{HS}. Once



(a) Complete schedule against actual observations for participant 002.



(b) After about 4 weeks the schedule is mature.

Figure 5: Maturity statistics for participant 002 (the maturity statistics for all 38 participants may be found in figures 14 to 23 of the appendix).

AP_0^{HS} has been identified, the homeset is constructed by including in HS any other access point that appears in a Wi-Fi scan together with AP_0^{HS}. This approach significantly increases the reliability of the homeset algorithm.

To measure this increase in reliability we define a metric called *stability*. We compute the stability π_x of an access point x over a time interval T_π, which we set to be at night between $3am$ and $4am$. If AP_0^{HS} is seen at least once within T_π, then it is reasonable to assume that the household must be occupied during the whole period. Indeed, although theoretically possible, it is unlikely that typical household occupants will leave the home between $3am$ and $4am$. However, in some scans registered in the period T_π AP_0^{HS} does not appear. If the homeset algorithm relied on AP_0^{HS} only, the household would be declared as occupied in given slots within the period T_π and unoccupied in others. This instability would clearly cause false negatives to appear and thus decrease the reliability of our algorithm. To demonstrate that the homeset approach significantly improves on this aspect, we thus compute the stability π_x as the ratio of two quantities. The numerator is the total number of scans in which the access point x appears in the period T_π. The denominator is the total number of scans in the period T_π, whereby the scans are counted only if the access point x is seen at least once in the period T_π. A value of π_x equal to 1 thus means that if the access point is seen on any given night, it is going to be seen in all other scans between $3am$ and $4am$ and thus that it is a stable indicator of household occupancy.

The rationale behind the homeset algorithm is that a set of access points has a higher stability than a single one, even if this one is the private access point of the household. Table 1 shows evidence of this observation for selected participants included in our MDC data set. For participant 009, for instance, using the whole HS instead of the single primary access point only, increased stability from 0.477 to 0.954.

3.2 Maturity of the schedules

A given probabilistic occupancy schedule can be considered mature only when sufficient data has been collected to construct it. In real settings, the actual maturity of the schedule must be measured before it starts being used to, for instance, control a heating system. We say that a schedule is *mature* when at least 95% of the slots contain at least 1 observation. For most of the participants in our MDC data set, maturity is achieved within 4 weeks.

The continuous straight line in figure 14(a) shows the total number of observations that would be counted if we had at least one Wi-Fi scan in each slot. The other two curves compare this "ideal" complete schedule with the average and median number of scans that are actually observed for an exemplary participant in our MDC data set. This plot clearly shows that during any week, the actual number of observation is smaller than the number of slots. However, if the user can be observed over several weeks, then maturity can be reached quickly, as shown in figure 19(a).

ID	$\pi_{AP_0^{HS}}$	$\omega_{AP_0^{HS}}$	π_{HS}	ω_{HS}	Score	In HS?
002	0.555	0.953	0.963	1.0	13	✓
005	0.229	0.424	0.414	0.909	2	?
007	0.859	0.654	0.892	0.946	16	✓
009	0.477	0.78	0.954	0.962	n.a.	n.a.
010	0.75	0.678	0.852	0.956	2	?
017	0.805	0.978	0.981	0.985	8	?
023	0.715	0.487	0.987	1.0	16	✓
026	0.588	0.875	0.963	0.971	3	?
034	0.883	0.481	0.866	0.57	16	✓
042	0.674	0.678	0.964	0.931	10	✓
050	0.668	0.948	0.979	1.0	10	✓
051	0.516	0.982	0.985	1.0	2	?
056	0.943	0.975	0.985	1.0	6	?
060	0.921	0.977	0.983	0.996	12	✓
063	0.851	1.0	0.995	1.0	5	?
068	0.857	0.912	0.998	1.0	5	?
075	0.634	0.481	0.892	0.659	16	✓
077	0.76	0.875	0.961	0.938	n.a.	n.a.
082	0.899	0.968	0.992	0.989	16	✓
083	0.664	0.988	0.998	1.0	16	✓
089	0.512	0.643	0.971	0.857	5	?
094	0.794	0.584	0.832	0.887	n.a.	n.a.
109	0.468	0.884	0.94	0.977	9	?
111	0.676	0.462	0.975	1.0	11	✓
117	0.375	0.825	0.964	0.997	15	✓
120	0.77	0.72	0.96	0.805	13	✓
123	0.813	1.0	0.994	1.0	5	?
126	0.546	0.841	0.957	0.955	2	?
127	0.704	0.689	0.949	0.974	16	✓
139	0.472	0.391	0.916	0.457	n.a.	n.a.
141	0.314	0.839	0.985	0.873	7	?
160	0.538	0.876	0.984	1.0	16	✓
165	0.615	0.968	0.962	0.992	n.a.	n.a.
169	0.692	0.736	0.98	1.0	14	✓
172	0.476	0.915	0.958	0.954	15	✓
179	0.812	0.368	0.971	0.974	16	✓
185	0.448	0.696	0.972	0.983	10	✓
186	0.914	1.0	0.964	1.0	16	✓

Table 1: Empirical probability ω and stability π of the primary access point A_0^{HS} only and the extended set of access points, the homeset HS, for all participants included in the data set (n.a.: not available, ?: score too low)

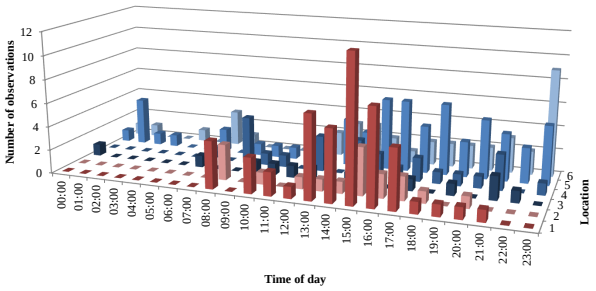


Figure 6: Time-frequency analysis of the anonymised locations for participant 002. Locations with less than 10 observations are excluded.

3.3 Validating the homeset algorithm

In order to thoroughly validate the homeset algorithm, a precise schedule of the absence from or presence in, the household of the mobile phone owners would be necessary. As this information is not available in the MDC data set, we set out to validate our findings indirectly by verifying whether the access points included in the homeset are plausibly close to the location of the participants' homes. To this end, we used the GPS data available in the MDC data set and considered the fact this data had been partially modified in order to protect the privacy of the participants. In particular, the latitude and longitude coordinates of sensitive places, like the participants' homes or workplaces, have been occasionally truncated to the 3rd decimal digit. As the coordinates are reported along with a timestamp, we could retrieve statistics about *when* participants were in sensitive places, even though it was not possible to retrieve *where* exactly the participants were at that specific time.

We thus first extract all the truncated instances of the GPS data from the data set. We then assign each unique pair of truncated latitude and longitude coordinates to a symbolic location k. For each location, we then create a frequency count vector $\vec{CV}_k = (c_0, c_1, \ldots, c_{23})$ with 24 elements, one for each hour of the day. Over the whole data set, we then count the number of occurrences of a location k in a given hour of the day and store this value in the corresponding element of the vector CV_k. We thus count how many times a specific symbolic location has been "anonymised".

Figure 6 shows the results of this analysis for participant 002, whereby we only display the 6 most relevant symbolic locations. As visible in this picture, location 1 is anonymised most of the times between $1pm$ and $5pm$ and is never anonymised before $8am$ or after $9pm$. We thus conjecture that this location corresponds to the workplace of the participant, as it is likely that between $1pm$ and $5pm$ the participant is at work and thus there is a higher need to truncate coordinates that correspond to this sensitive location. On the other side, location 5 is the one that is anonymised most frequently and consistently over the whole course of the day. Therefore, we conjecture that this is the location of the home of the participant.

In order to automatically assess if a particular set of coordinates can identify a home location, we compute a score for each location. To make results comparable, we round CV_k to binary values and multiply it with a weighting vector $\vec{w} = (w_0, w_1, \ldots, w_{23})$. Times between 9 and 17 (i.e. w_9 to w_{17}) are set to $\frac{2}{7}$ while all other times are set to 1. We chose this weighting assuming a normal nine to five schedule with little presence during the day except on weekends. A set of coordinates can score a maximum of 18.3 points under this metric. We have chosen a threshold of 10 for a location to be accepted as a possible home location.

Once we retrieved the (truncated and thus anonymised) location of the home of each participant using the method described above, we compare the symbolic location with the GPS coordinates of the Wi-Fi access points. To this end, we compute the locations of the access points using temporal matching between the Wi-Fi and anonymised GPS data. For 20 out of the 38 participants included in the data set, a match was found. Of the remaining cases, 13 times the score of the candidate locations was below 10 and in 5 cases no anonymised coordinates could be found for the homeset access points. By comparing the homesets we could further identify 4 out of the 13 participants with

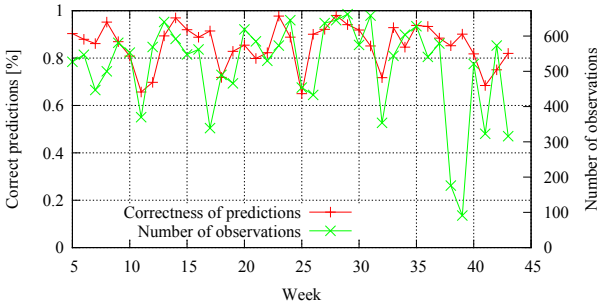


Figure 7: Prediction errors and observations for participant 060.

low scores as couples (i.e. intersecting homeset, similar schedule, similar age, male and female). As their candidate anonymised GPS locations were also identical, we could thus lift their combined score over the threshold and validate 4 additional participants. Thus, for the majority of the participants in the data set, we could verify that the coordinates of the symbolic location identified as the home of the participants corresponded with the coordinates of access points included in the homeset, thus establishing the reliability of our homeset algorithm.

4 Examples of the use of probabilistic occupancy schedules

The probabilistic schedules computed with the homeset algorithm can be used, among other things, to control home automation systems. In particular, these schedules can be used to adapt to actual schedules and *predict* the occupants' future behaviour in order to control the heating system with the goal of saving energy without losing comfort. In order to test the validity of this assumption we performed a preliminary evaluation. We used the first four weeks of data to learn the probabilistic schedules using the homeset algorithm. We then compared the probabilistic schedule with the observed participant occupancy in the subsequent weeks. We refer to a difference between these schedules as the *prediction error*.

Figure 7 exemplarily shows the results of this analysis for participant 060 – results for all other participants may found in figures 24 to 28 of the appendix. In this case, our probabilistic schedule correctly predicts the home occupancy of the participant for 80% of the observations.

This means that 80% of the observed slots over the course of a particular week have been classified correctly. This accuracy figure is only meaningful if the number of observations is close to the maximum (i.e. 672 15-minute time slots). In the case of participant 060, data for more than 80% of the slots are available per week, most of the time. However, during weeks 38 and 39, the fact that we have data for less than 200 slots each (e.g. less than 3 days of data) does not enable us to conclude that the prediction would produce meaningful results for these two weeks even though the accuracy remains high.

In order to quantify the prediction error of the final probabilistic schedule we compute a vector of the expected errors $\vec{m}$, where $t = 15$ is the length of the interval in minutes. This essentially is the probability of erring if we set the threshold for occupancy to 0.5.

$$\vec{m} = (m_0, \ldots, m_6) \textbf{ where } m_i = \sum_{j=0}^{n} min(p_{ij}, 1 - p_{ij}) \times t$$

Figure 8 shows the weekday expected errors $E(m_0, \ldots, m_4)$ plotted against the mean of the weekday absences for participants studying or working full-time. Like the expected errors, the weekday absences were obtained from the probabilistic schedule by regarding the participant as absent if $p_{ij} < 0.5$ and taking the longest continuous stretch of absence. It can be seen that participants working full-time are in general staying away from home longer and have a more regular schedule. Some exceptions are highlighted in the graph. Students are much more difficult to classify, however. For some, the actual schedules vary so much (i.e. p_{ij} is close to 0.5) that the probabilistic schedule

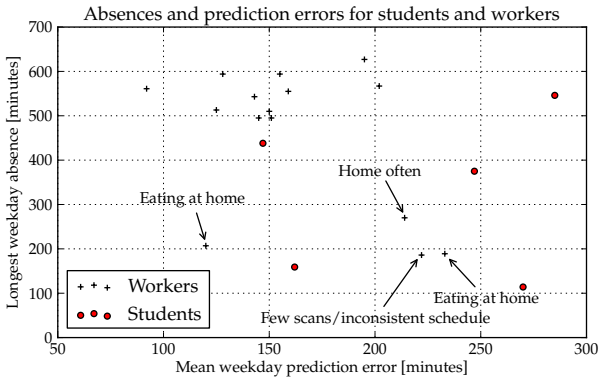


Figure 8: Weekday absences and prediction errors for full-time workers (crosses) and students (dots). Full time workers are away for longer continuous stretches of time and are more predictable.

cannot accurately determine for how long they can expected to be absent. This means that a prediction algorithm based merely on static schedules [5] would not consistently produce correct predictions. Therefore, to overcome the irregularity, it must identify the type of household and adapt to its occupants. This may mean either lowering the threshold (choosing comfort over savings) or performing next place prediction.

5 Conclusions

This report introduced a novel algorithm to compute occupancy schedules of private households using mobile phone data. Our *homeset* algorithm uses Wi-Fi scan data and is able to determine when a schedule is *mature* enough and can thus be used to, e.g. predict occupants' future behaviour. To validate our findings we developed a second technique that automatically validates the homeset using the anonymised GPS data contained in the MDC data set.

References

[1] D. Ashbrook and T. Starner. Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users. *Personal Ubiquitous Computing*, 7(5):275–286, October 2003.

[2] I. Constandache, S. Gaonkar, M. Sayler, R. Choudhury, and L. Cox. Enloc: Energy-efficient localization for mobile phones. In *Proc. of INFOCOM'09*. IEEE, March 2009.

[3] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding Individual Human Mobility Patterns. *Nature*, 453:779–782, June 2008.

[4] D. H. Kim, J. Hightower, R. Govindan, and D. Estrin. Discovering Semantically Meaningful Places from Pervasive RF-beacons. In *Proc. of UbiComp'09*. ACM, October 2009.

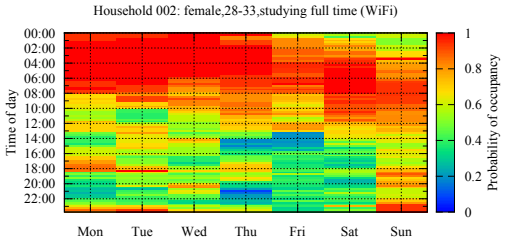
[5] J. Krumm and A. J. B. Brush. Learning Time-based Presence Probabilities. In *Proc. of Pervasive'11*. Springer, June 2011.

[6] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T. Do, O. Dousse, J. Eberle, and M. Miettinen. The Mobile Data Challenge: Big Data for Mobile Computing Research. In *Proc. of the MDC by Nokia Workshop, in conjunction with Pervasive'12*. Springer, June 2012.

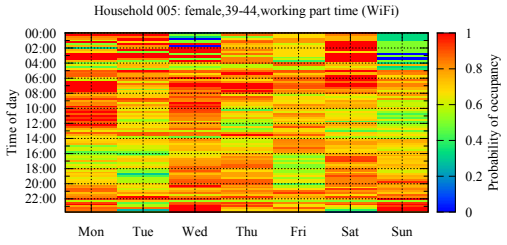
[7] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell. NextPlace: A Spatio-Temporal Prediction Framework for Pervasive Systems. In *Proc. of Pervasive'11*. Springer, June 2011.

[8] J. Scott, A. J. B. Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, and N. Villar. PreHeat: Controlling Home Heating Using Occupancy Prediction. In *Proc. of UbiComp'11*. ACM, September 2011.

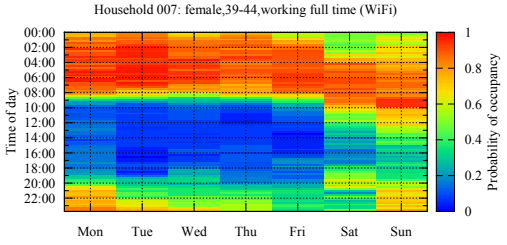
Appendix



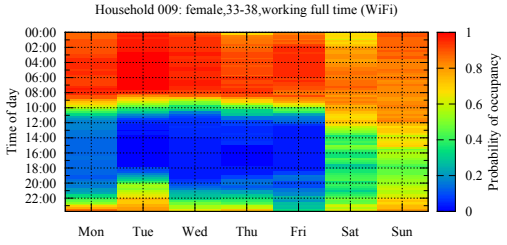
(a) Probabilistic schedule for participant 002.



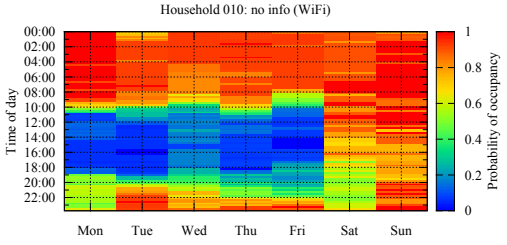
(b) Probabilistic schedule for participant 005.



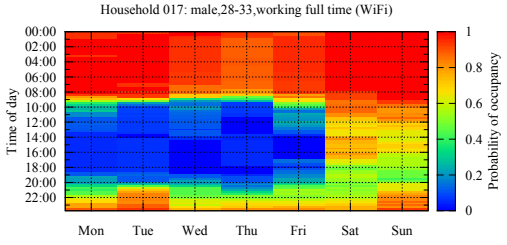
(c) Probabilistic schedule for participant 007.



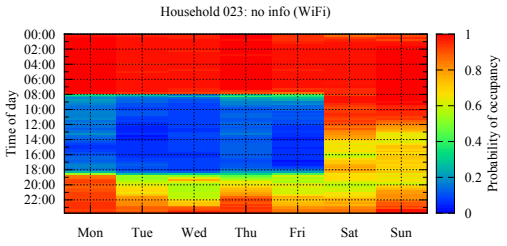
(d) Probabilistic schedule for participant 009.



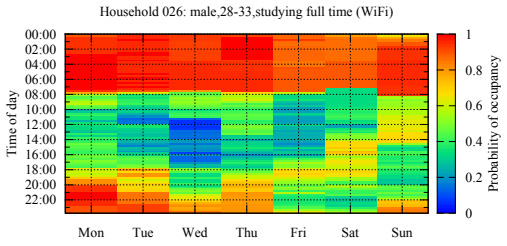
(e) Probabilistic schedule for participant 010.



(f) Probabilistic schedule for participant 017.

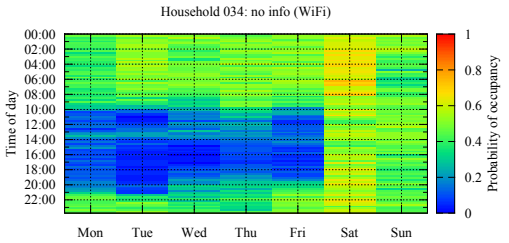


(g) Probabilistic schedule for participant 023.

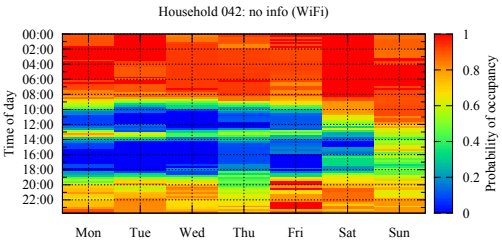


(h) Probabilistic schedule for participant 026.

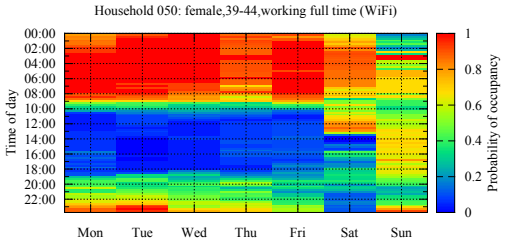
Figure 9: Probabilistic schedules for participants 002 to 026.



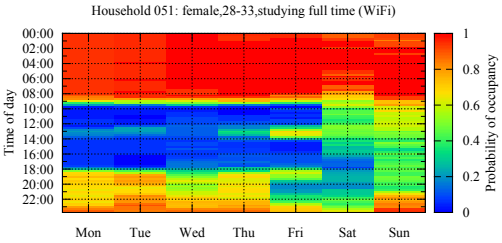
(a) Probabilistic schedule for participant 034.



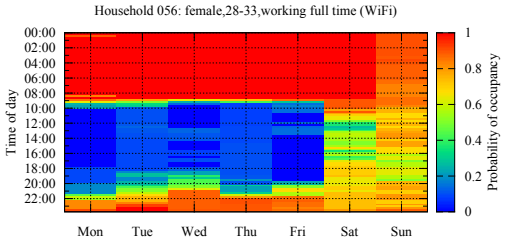
(b) Probabilistic schedule for participant 042.



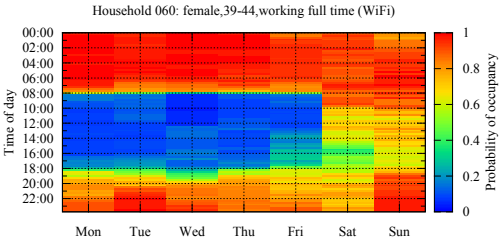
(c) Probabilistic schedule for participant 050.



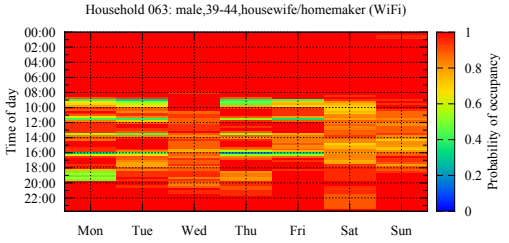
(d) Probabilistic schedule for participant 051.



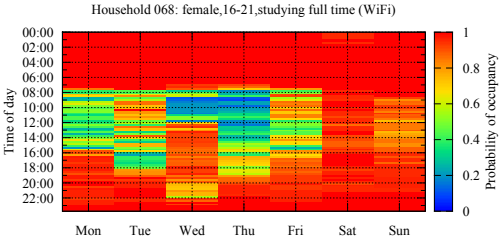
(e) Probabilistic schedule for participant 056.



(f) Probabilistic schedule for participant 060.

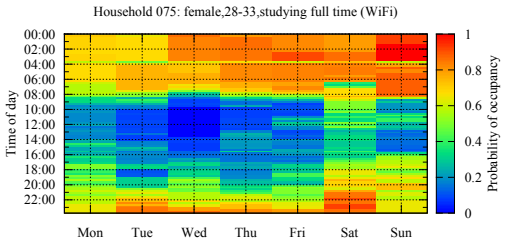


(g) Probabilistic schedule for participant 063.

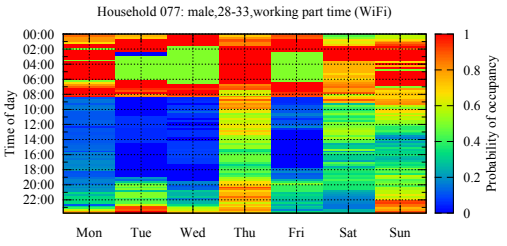


(h) Probabilistic schedule for participant 068.

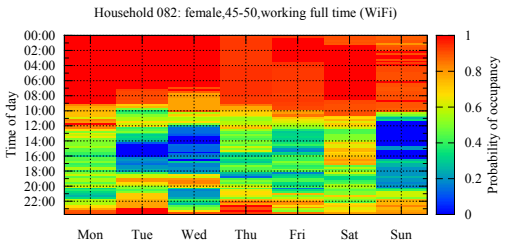
Figure 10: Probabilistic schedules for participants 034 to 068.



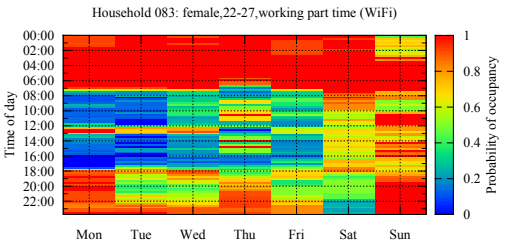
(a) Probabilistic schedule for participant 075.



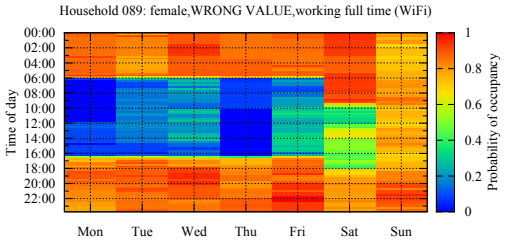
(b) Probabilistic schedule for participant 077.



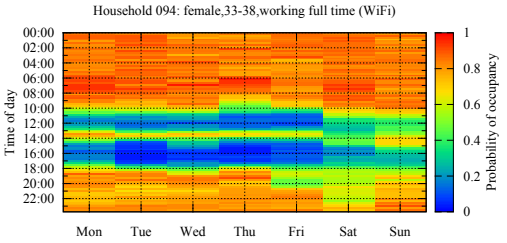
(c) Probabilistic schedule for participant 082.



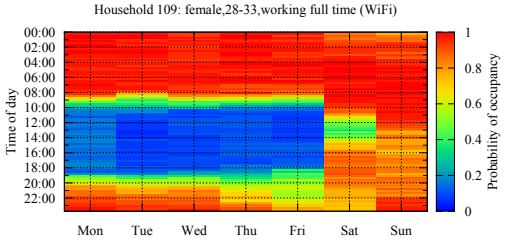
(d) Probabilistic schedule for participant 083.



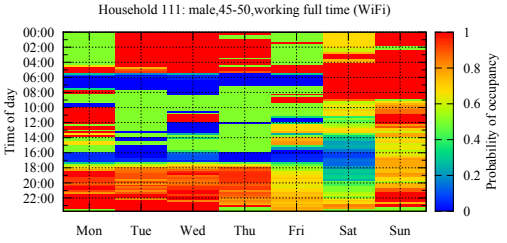
(e) Probabilistic schedule for participant 089.



(f) Probabilistic schedule for participant 094.

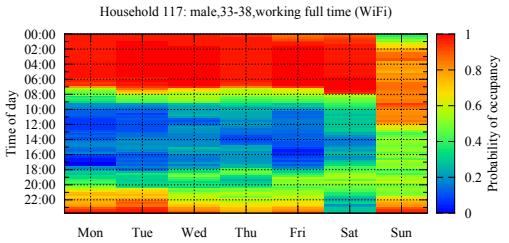


(g) Probabilistic schedule for participant 109.

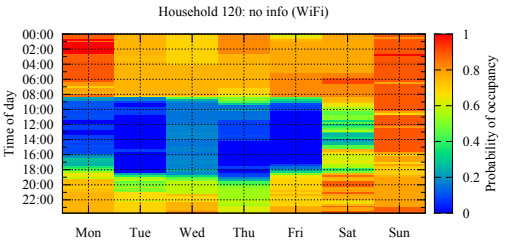


(h) Probabilistic schedule for participant 111.

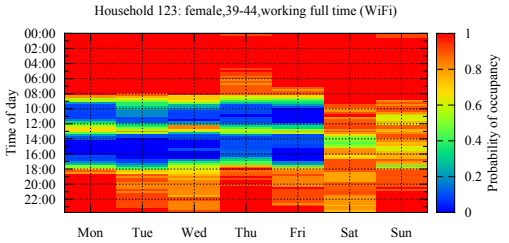
Figure 11: Probabilistic schedules for participants 075 to 111.



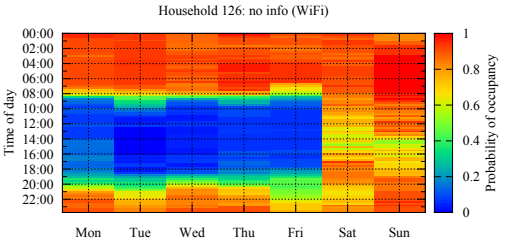
(a) Probabilistic schedule for participant 117.



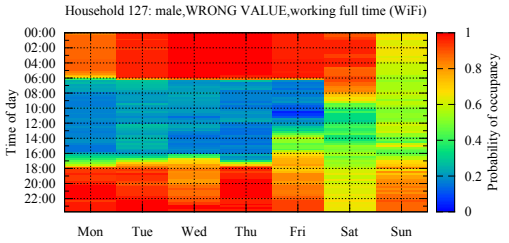
(b) Probabilistic schedule for participant 120.



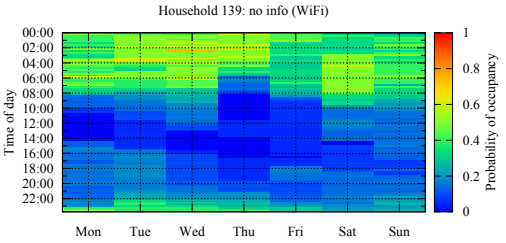
(c) Probabilistic schedule for participant 123.



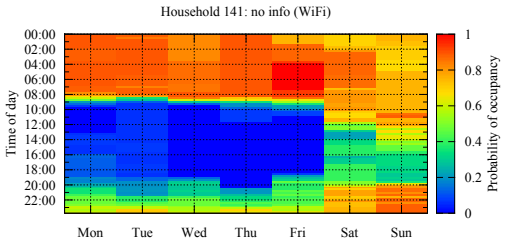
(d) Probabilistic schedule for participant 126.



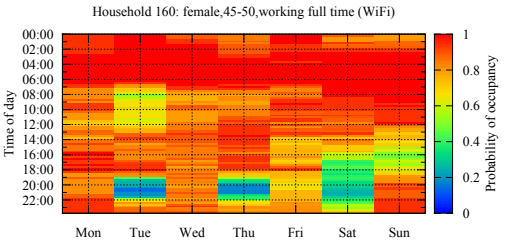
(e) Probabilistic schedule for participant 127.



(f) Probabilistic schedule for participant 139.

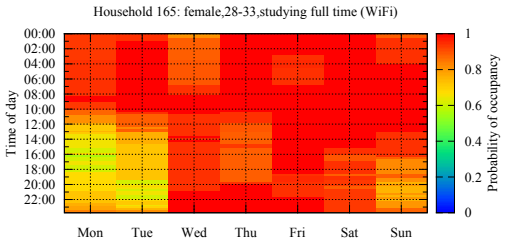


(g) Probabilistic schedule for participant 141.

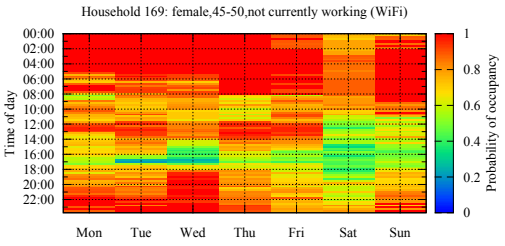


(h) Probabilistic schedule for participant 160.

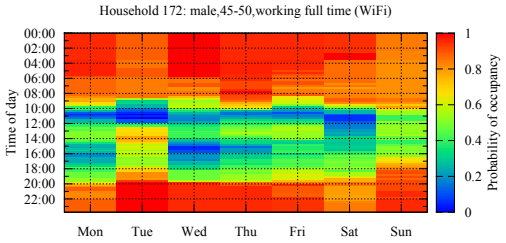
Figure 12: Probabilistic schedules for participants 117 to 160.



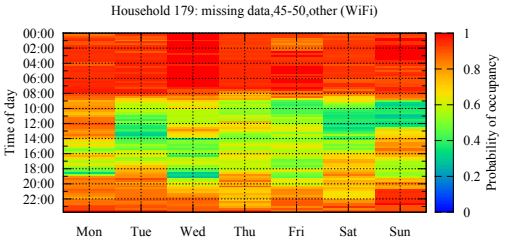
(a) Probabilistic schedule for participant 165.



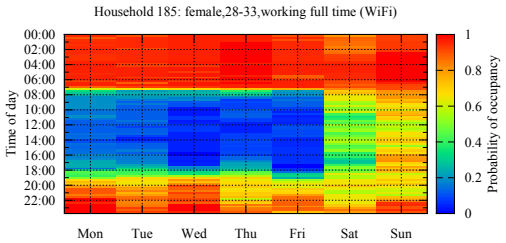
(b) Probabilistic schedule for participant 169.



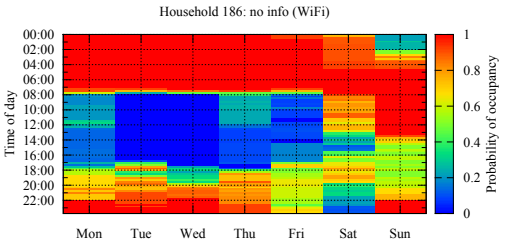
(c) Probabilistic schedule for participant 172.



(d) Probabilistic schedule for participant 179.

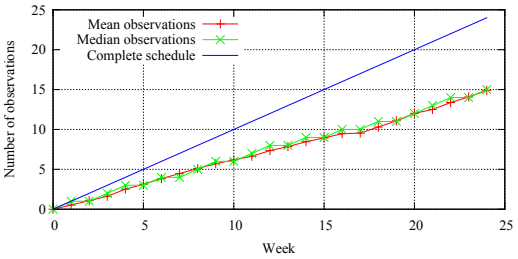


(e) Probabilistic schedule for participant 185.

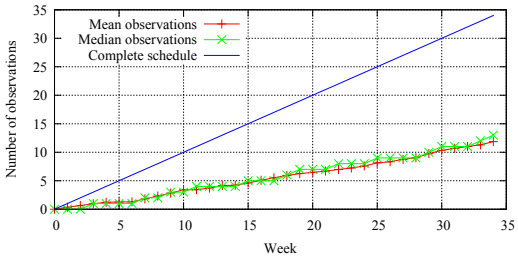


(f) Probabilistic schedule for participant 186.

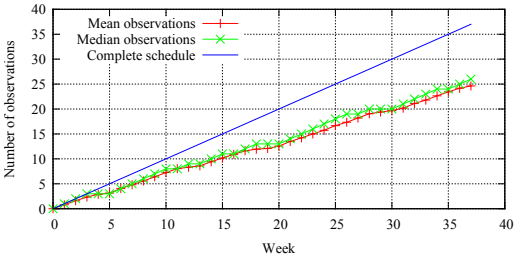
Figure 13: Probabilistic schedules for participants 165 to 186.



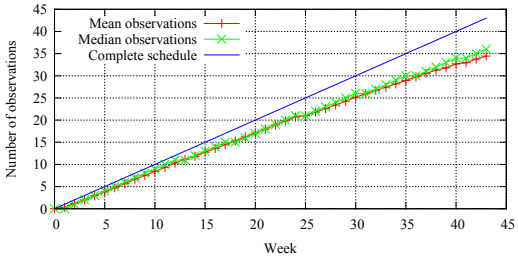
(a) Observation counts for participant 002.



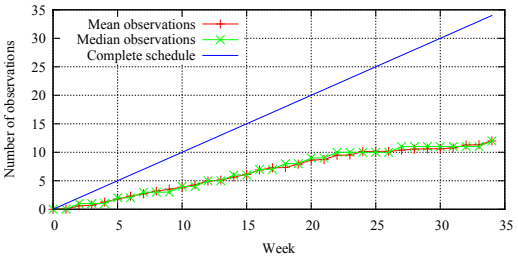
(b) Observation counts for participant 005.



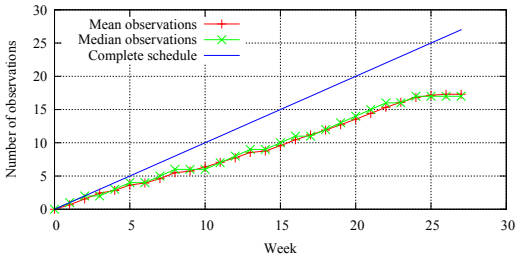
(c) Observation counts for participant 007.



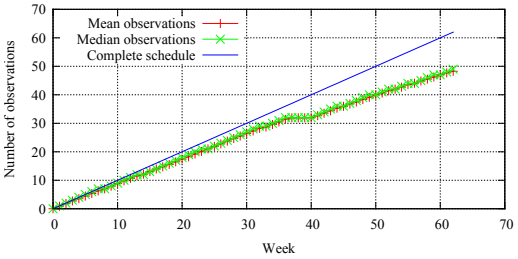
(d) Observation counts for participant 009.



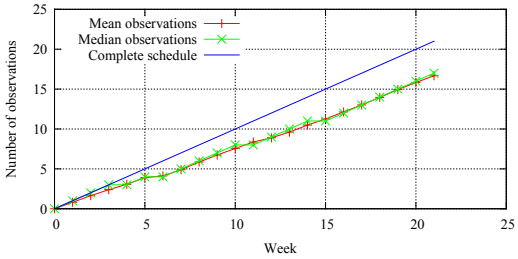
(e) Observation counts for participant 010.



(f) Observation counts for participant 017.

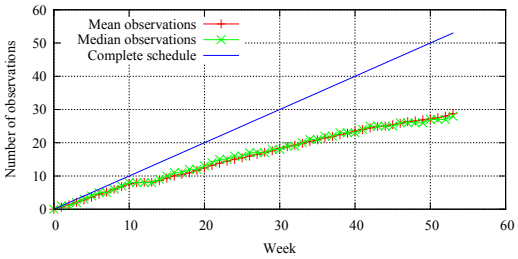


(g) Observation counts for participant 023.

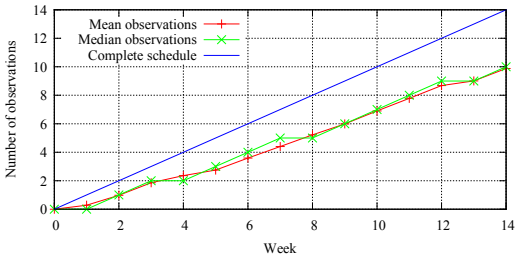


(h) Observation counts for participant 026.

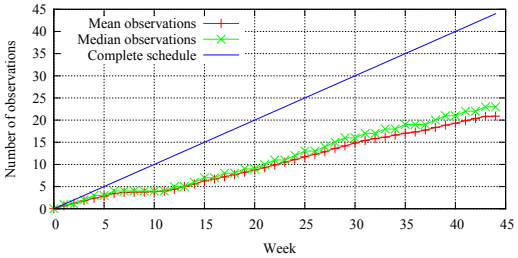
Figure 14: Observation counts for participants 002 to 026.



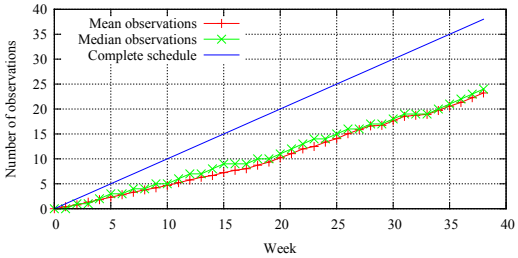
(a) Observation counts for participant 034.



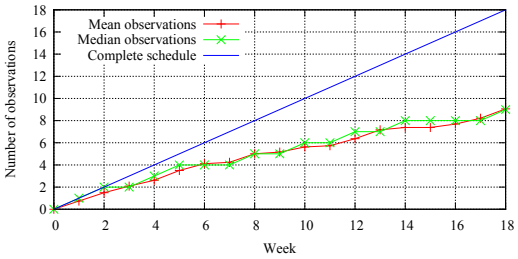
(b) Observation counts for participant 042.



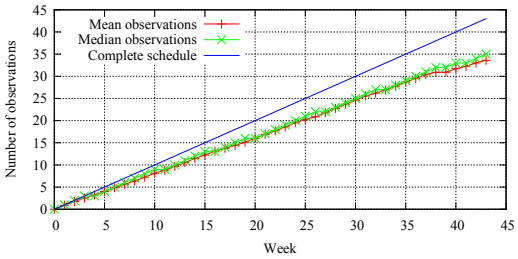
(c) Observation counts for participant 050.



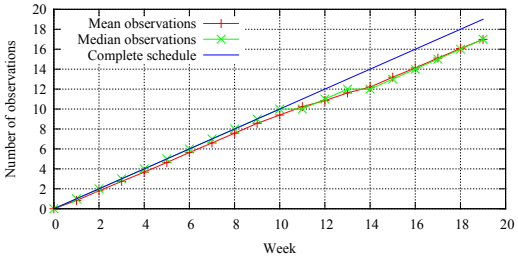
(d) Observation counts for participant 051.



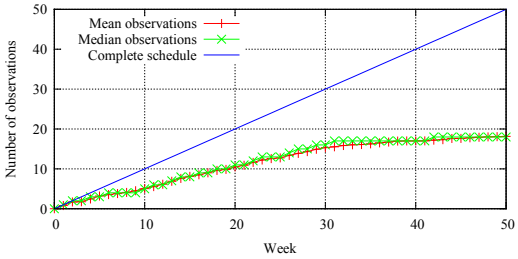
(e) Observation counts for participant 056.



(f) Observation counts for participant 060.

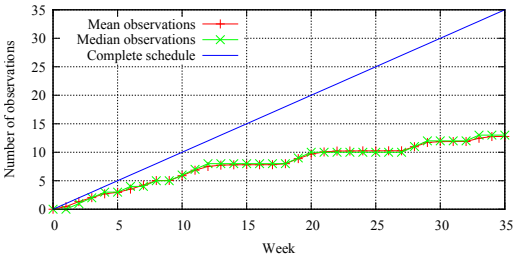


(g) Observation counts for participant 063.

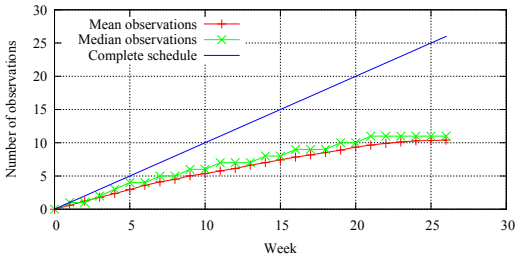


(h) Observation counts for participant 068.

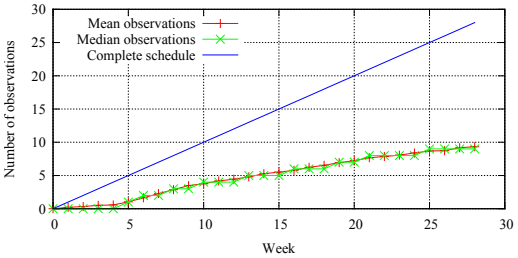
Figure 15: Observation counts for participants 034 to 068.



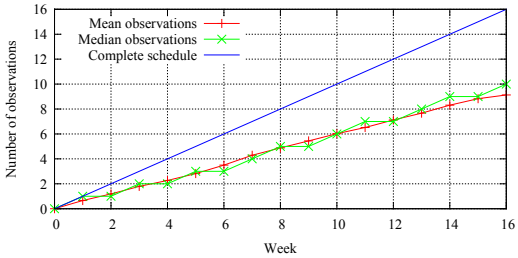
(a) Observation counts for participant 075.



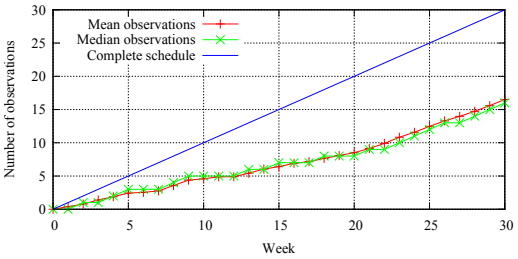
(b) Observation counts for participant 077.



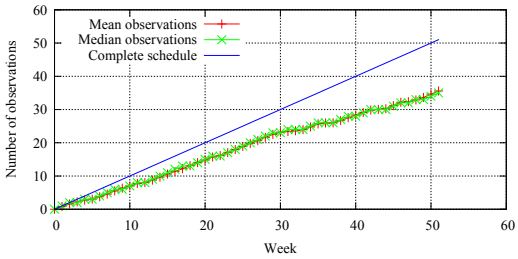
(c) Observation counts for participant 082.



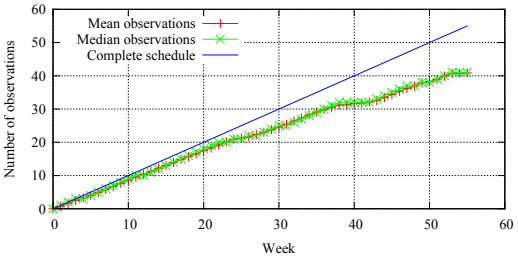
(d) Observation counts for participant 083.



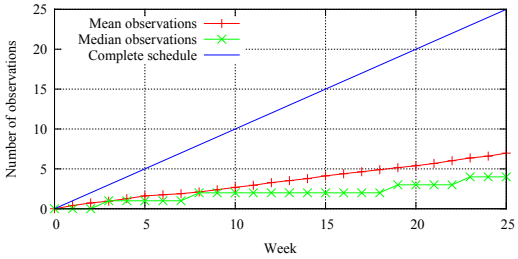
(e) Observation counts for participant 089.



(f) Observation counts for participant 094.

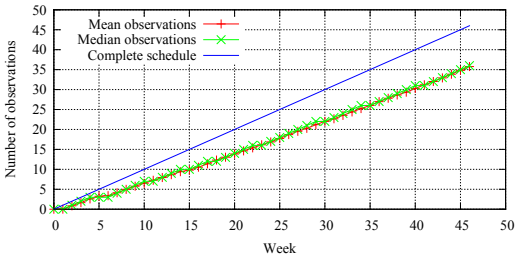


(g) Observation counts for participant 109.

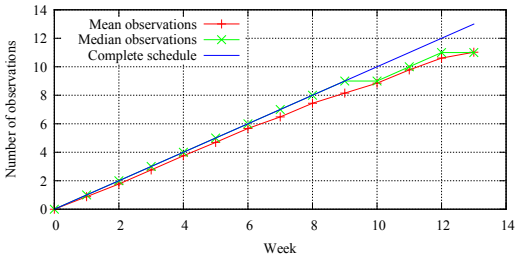


(h) Observation counts for participant 111.

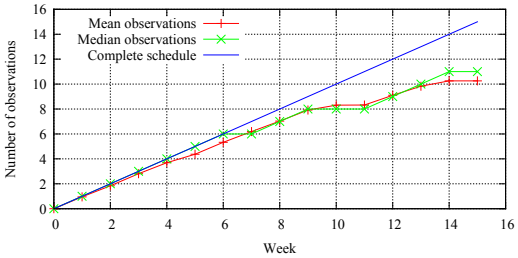
Figure 16: Observation counts for participants 075 to 111.



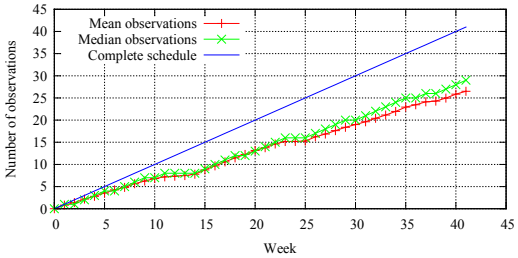
(a) Observation counts for participant 117.



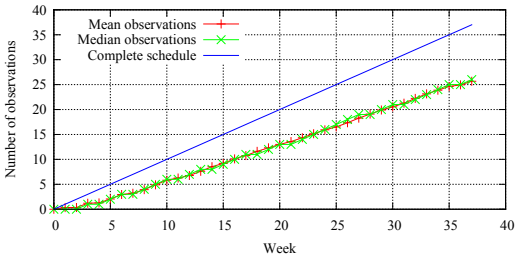
(b) Observation counts for participant 120.



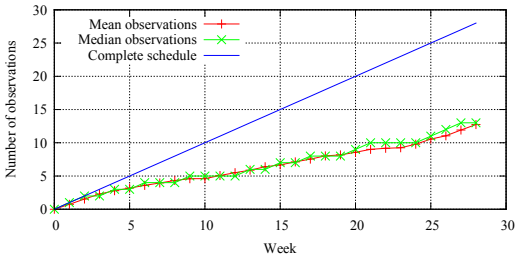
(c) Observation counts for participant 123.



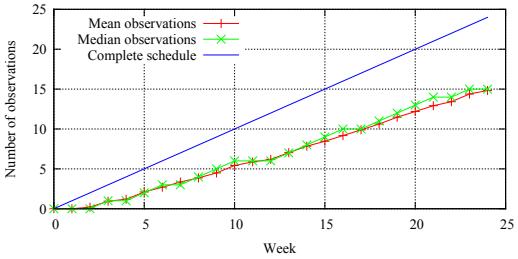
(d) Observation counts for participant 126.



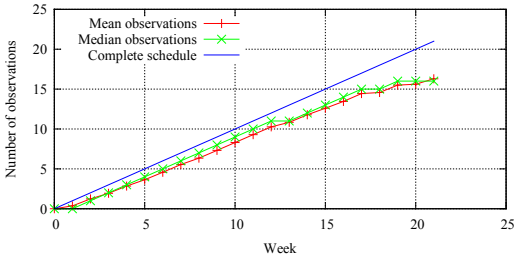
(e) Observation counts for participant 127.



(f) Observation counts for participant 139.

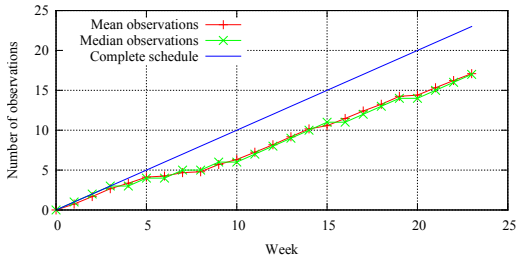


(g) Observation counts for participant 141.

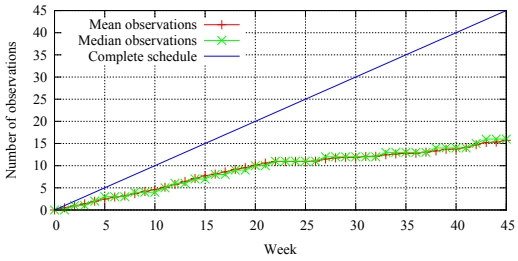


(h) Observation counts for participant 160.

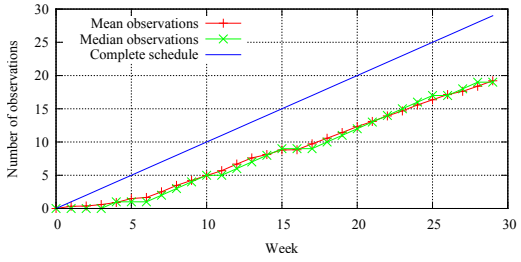
Figure 17: Observation counts for participants 117 to 160.



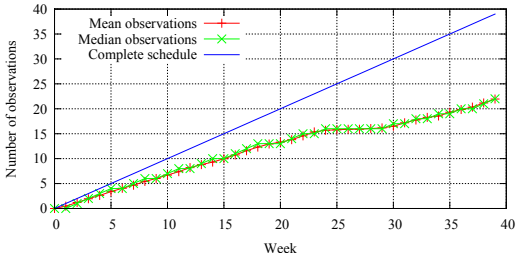
(a) Observation counts for participant 165.



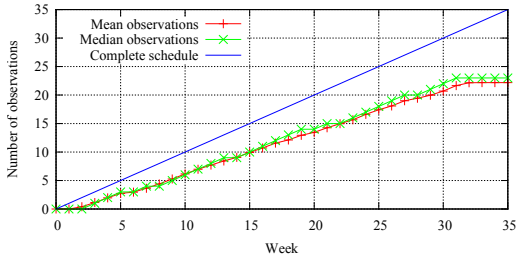
(b) Observation counts for participant 169.



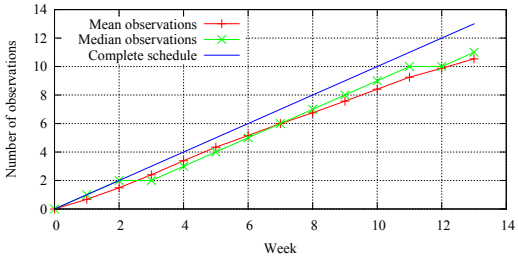
(c) Observation counts for participant 172.



(d) Observation counts for participant 179.

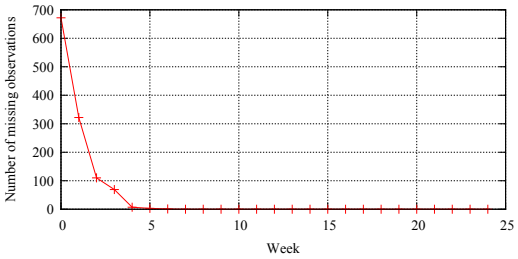


(e) Observation counts for participant 185.

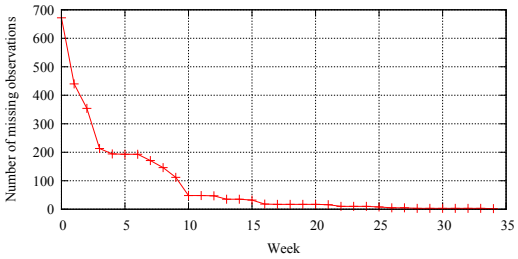


(f) Observation counts for participant 186.

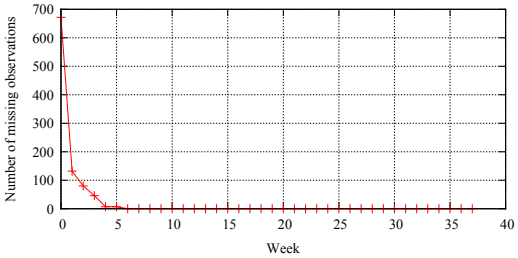
Figure 18: Observation counts for participants 165 to 186.



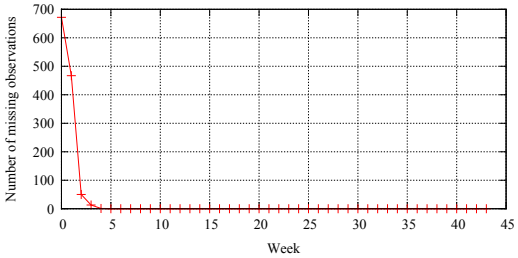
(a) Missing observations for participant 002.



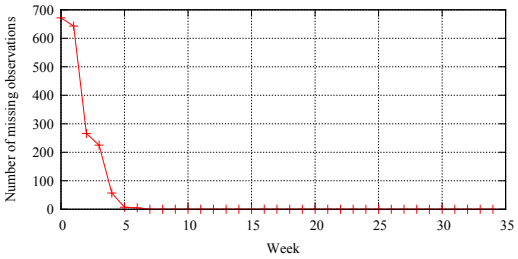
(b) Missing observations for participant 005.



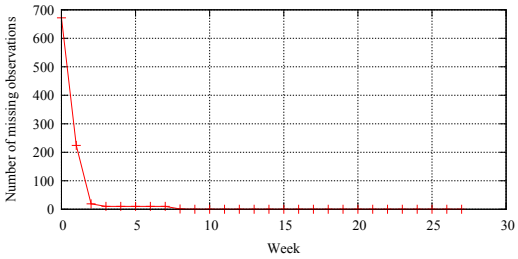
(c) Missing observations for participant 007.



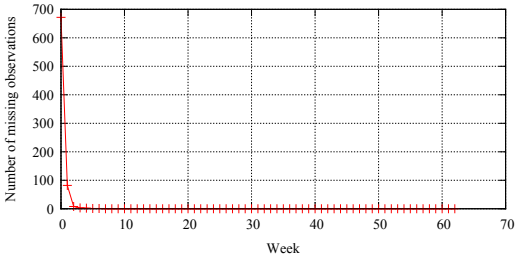
(d) Missing observations for participant 009.



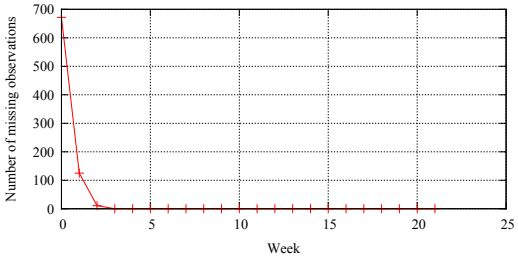
(e) Missing observations for participant 010.



(f) Missing observations for participant 017.

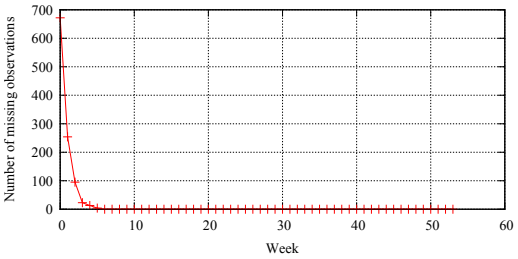


(g) Missing observations for participant 023.

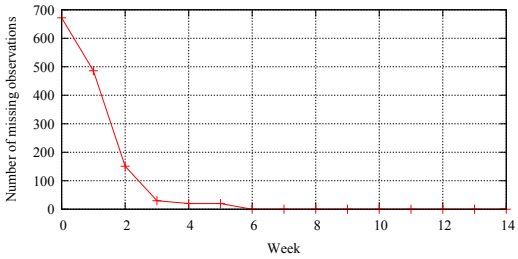


(h) Missing observations for participant 026.

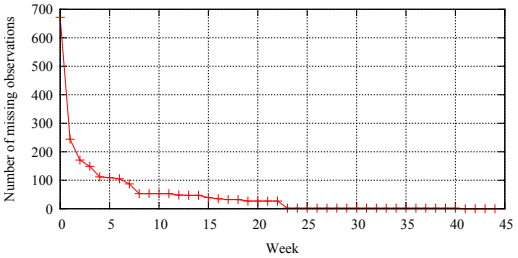
Figure 19: Missing observations for participants 002 to 026.



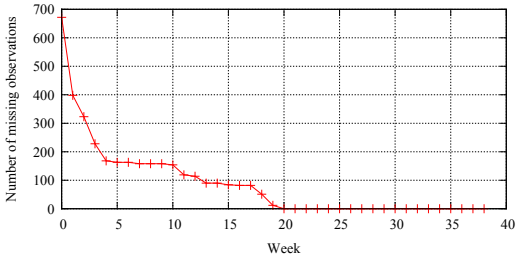
(a) Missing observations for participant 034.



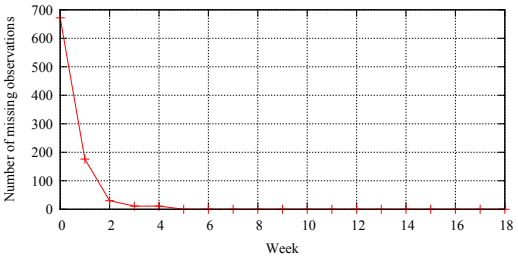
(b) Missing observations for participant 042.



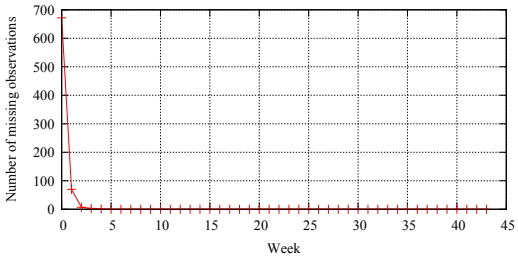
(c) Missing observations for participant 050.



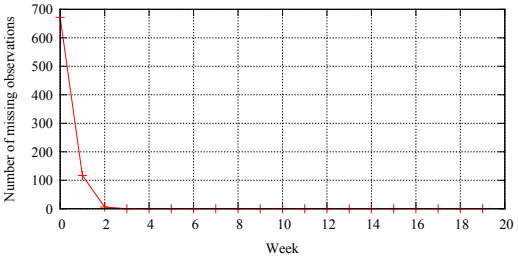
(d) Missing observations for participant 051.



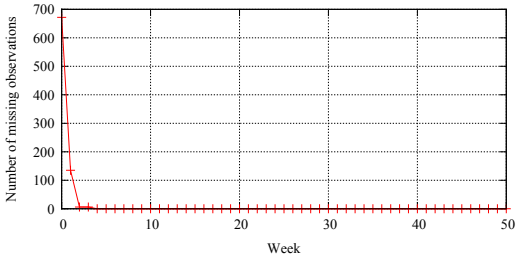
(e) Missing observations for participant 056.



(f) Missing observations for participant 060.

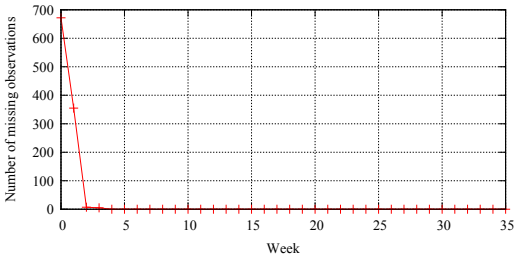


(g) Missing observations for participant 063.

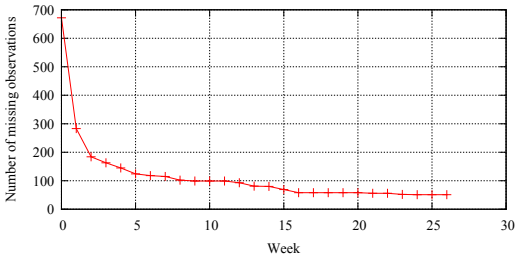


(h) Missing observations for participant 068.

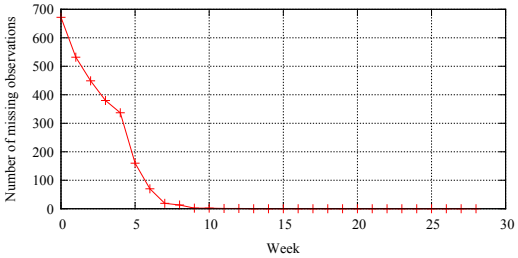
Figure 20: Missing observations for participants 034 to 068.



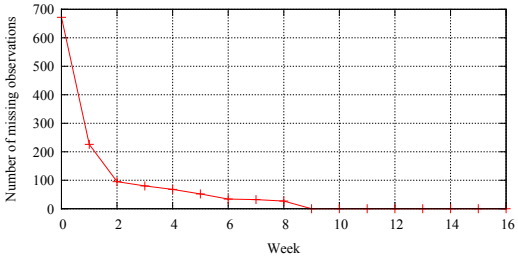
(a) Missing observations for participant 075.



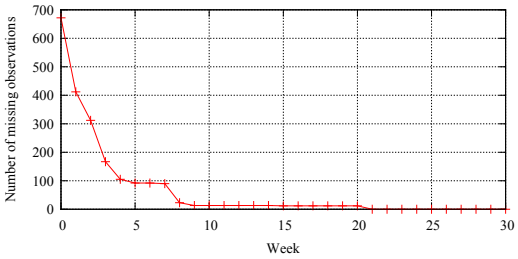
(b) Missing observations for participant 077.



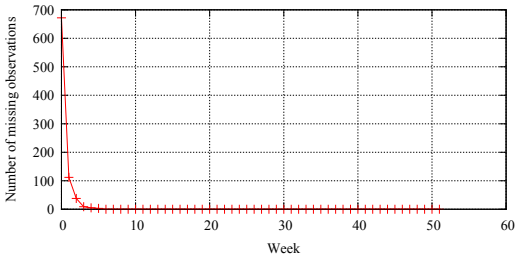
(c) Missing observations for participant 082.



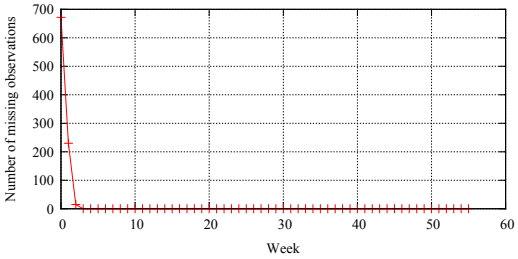
(d) Missing observations for participant 083.



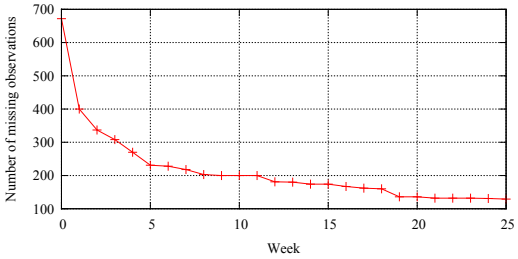
(e) Missing observations for participant 089.



(f) Missing observations for participant 094.

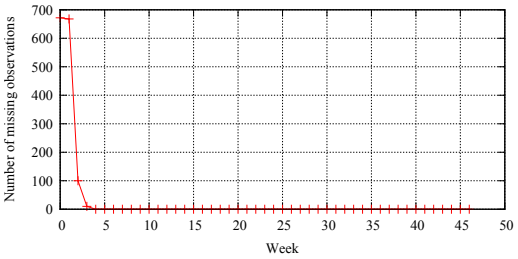


(g) Missing observations for participant 109.

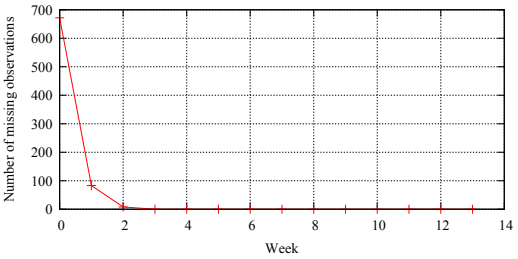


(h) Missing observations for participant 111.

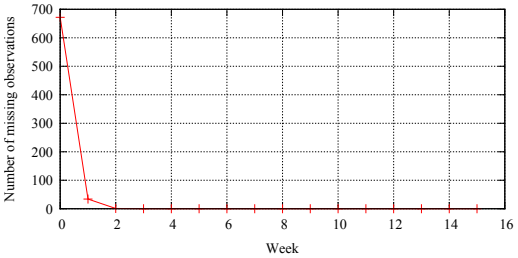
Figure 21: Missing observations for participants 075 to 111.



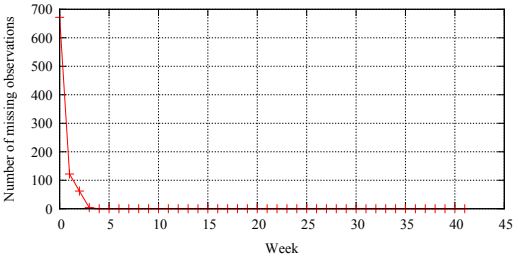
(a) Missing observations for participant 117.



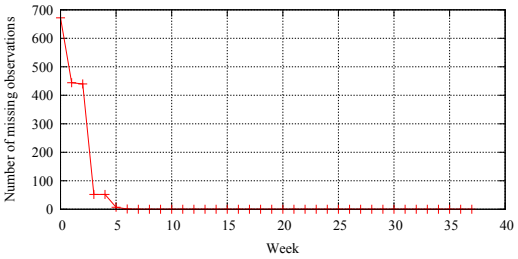
(b) Missing observations for participant 120.



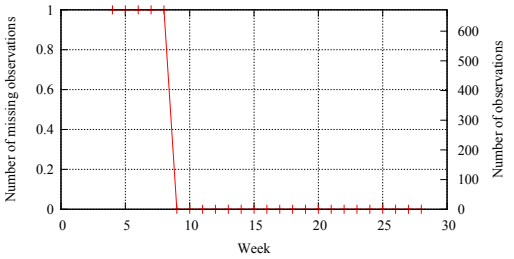
(c) Missing observations for participant 123.



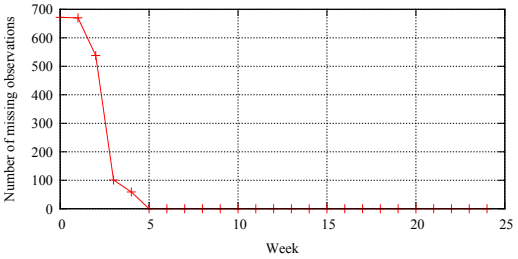
(d) Missing observations for participant 126.



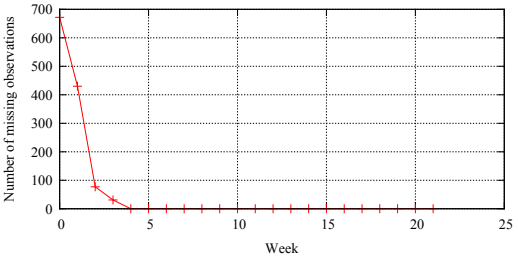
(e) Missing observations for participant 127.



(f) Missing observations for participant 139.

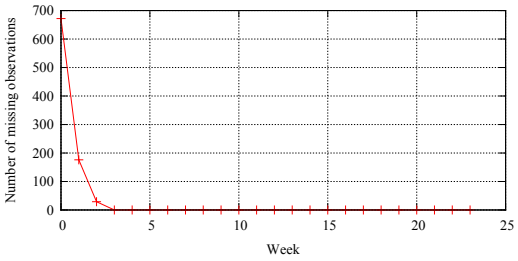


(g) Missing observations for participant 141.

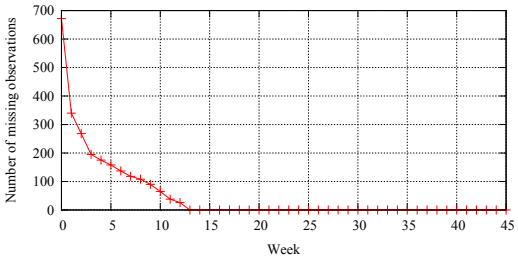


(h) Missing observations for participant 160.

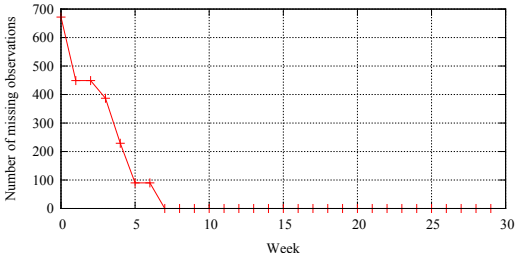
Figure 22: Missing observations for participants 117 to 160.



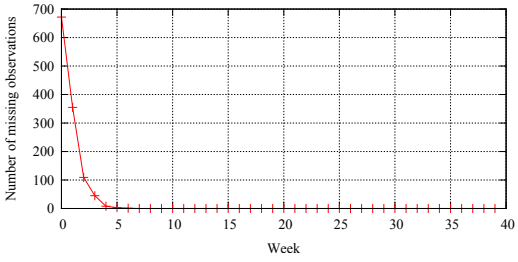
(a) Missing observations for participant 165.



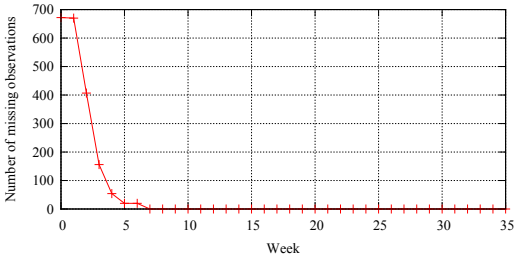
(b) Missing observations for participant 169.



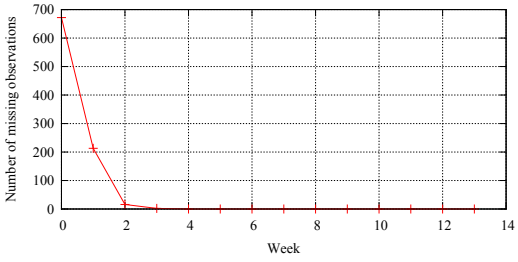
(c) Missing observations for participant 172.



(d) Missing observations for participant 179.

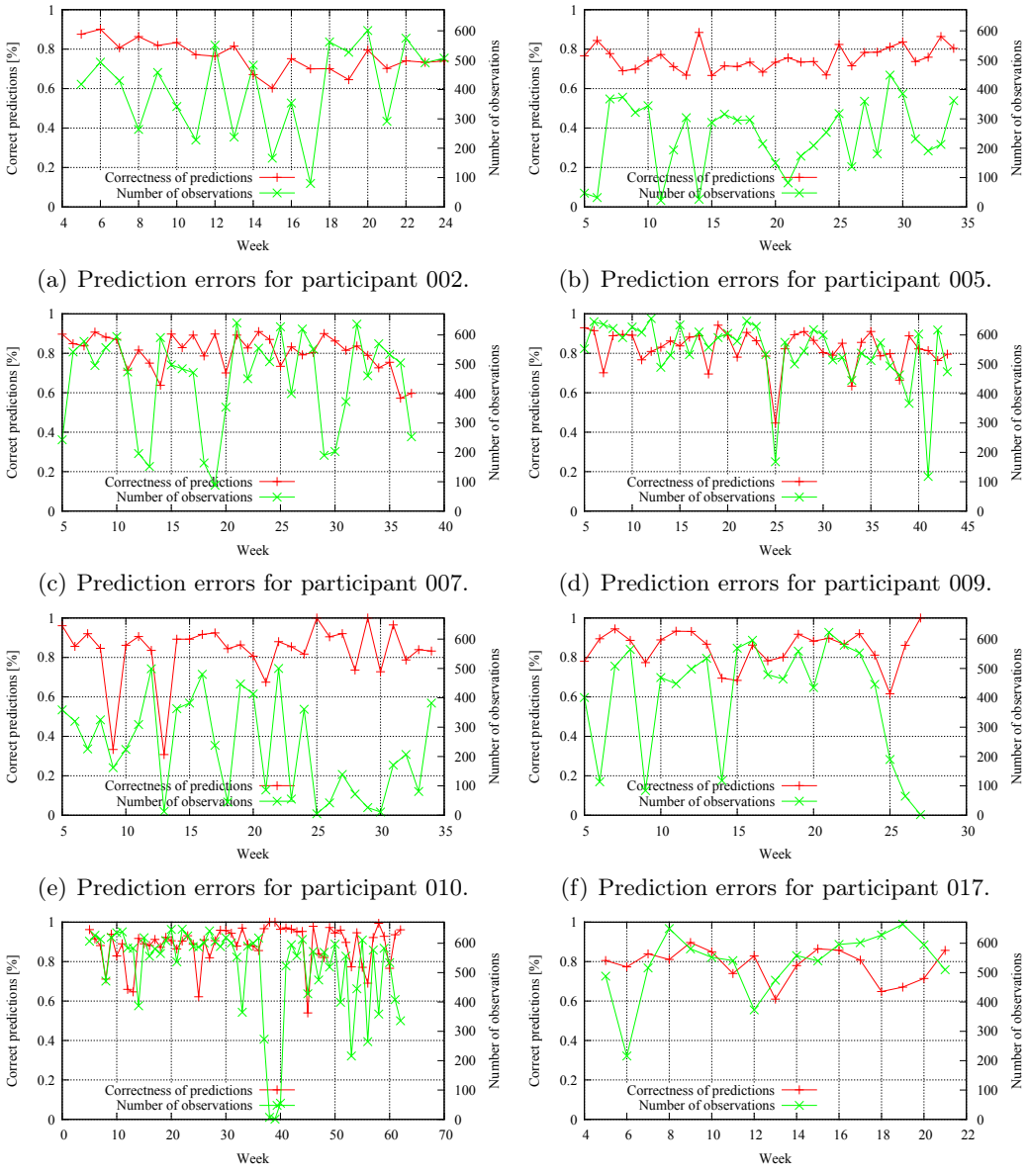


(e) Missing observations for participant 185.



(f) Missing observations for participant 186.

Figure 23: Missing observations for participants 165 to 186.



(a) Prediction errors for participant 002.

(b) Prediction errors for participant 005.

(c) Prediction errors for participant 007.

(d) Prediction errors for participant 009.

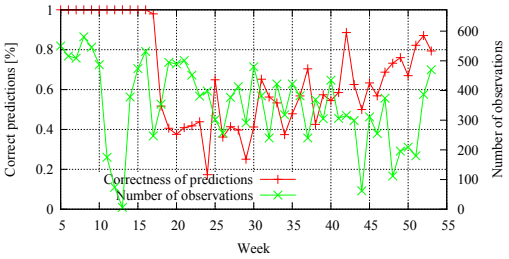
(e) Prediction errors for participant 010.

(f) Prediction errors for participant 017.

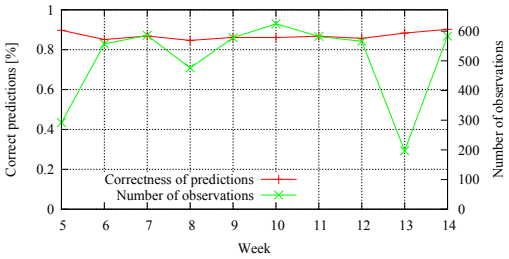
(g) Prediction errors for participant 023.

(h) Prediction errors for participant 026.

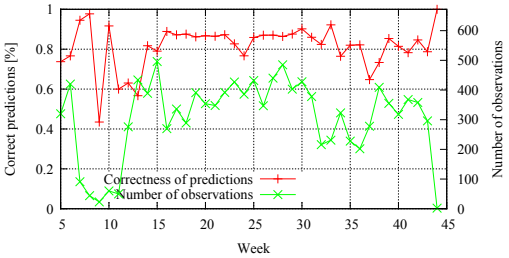
Figure 24: Prediction errors and number of observations for participants 002 to 026.



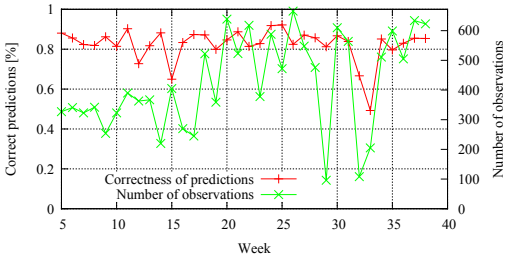
(a) Prediction errors for participant 034.



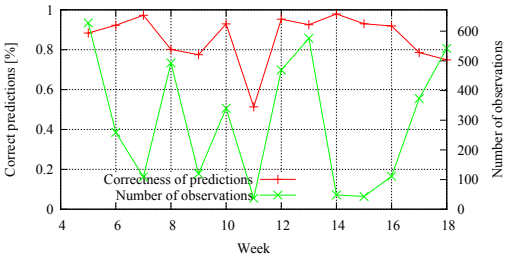
(b) Prediction errors for participant 042.



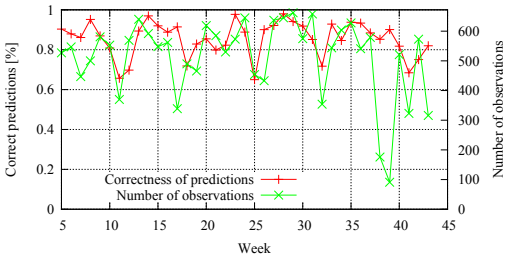
(c) Prediction errors for participant 050.



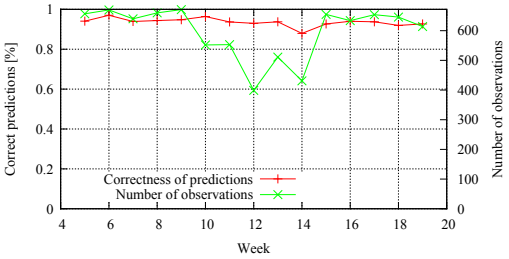
(d) Prediction errors for participant 051.



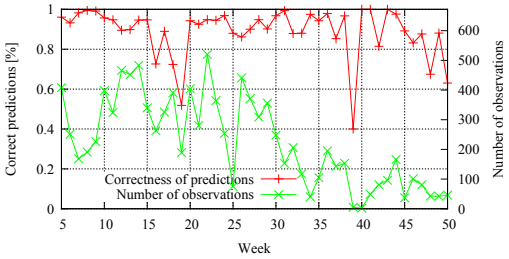
(e) Prediction errors for participant 056.



(f) Prediction errors for participant 060.

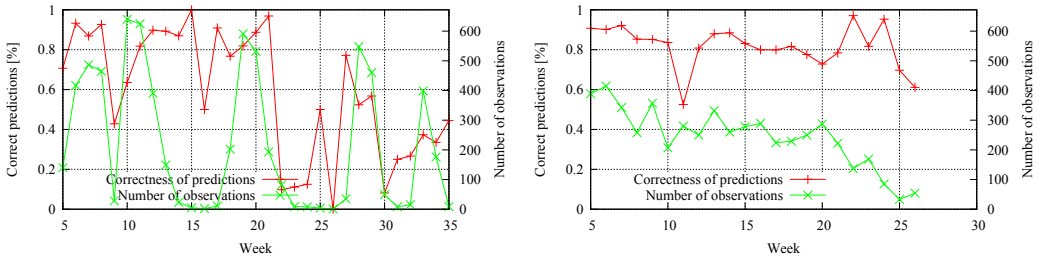


(g) Prediction errors for participant 063.



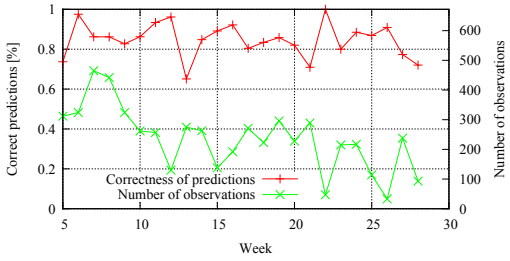
(h) Prediction errors for participant 068.

Figure 25: Prediction errors and number of observations for participants 034 to 068.

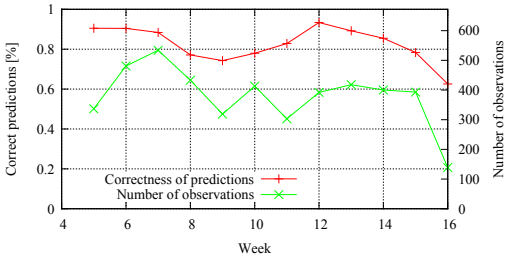


(a) Prediction errors for participant 075.

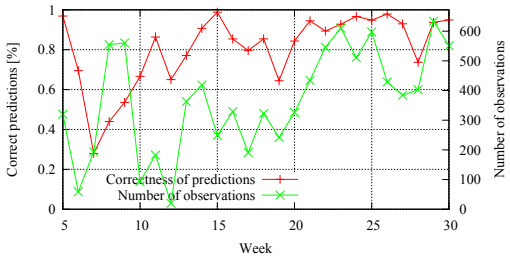
(b) Prediction errors for participant 077.



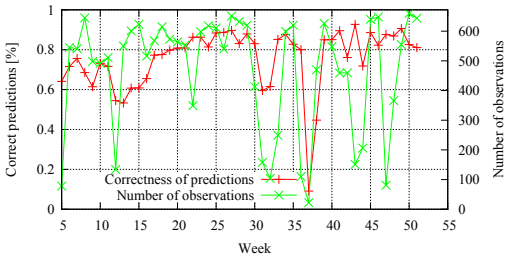
(c) Prediction errors for participant 082.



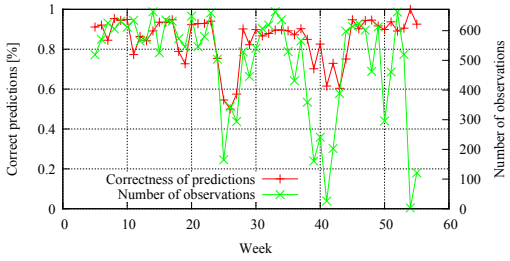
(d) Prediction errors for participant 083.



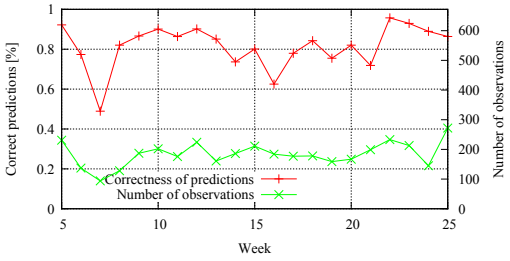
(e) Prediction errors for participant 089.



(f) Prediction errors for participant 094.

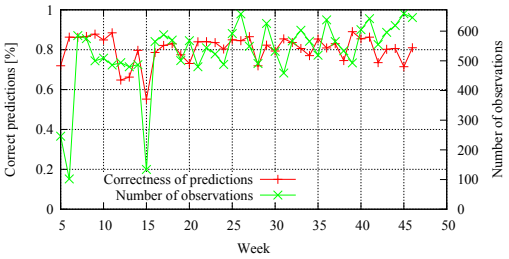


(g) Prediction errors for participant 109.

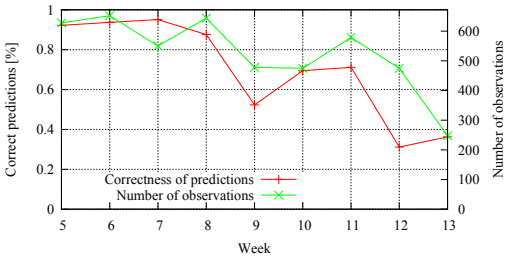


(h) Prediction errors for participant 111.

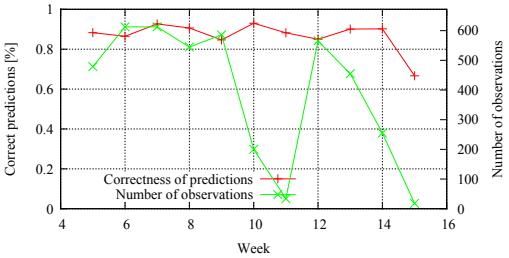
Figure 26: Prediction errors and number of observations for participants 075 to 111.



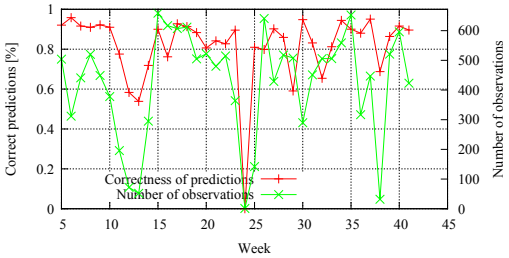
(a) Prediction errors for participant 117.



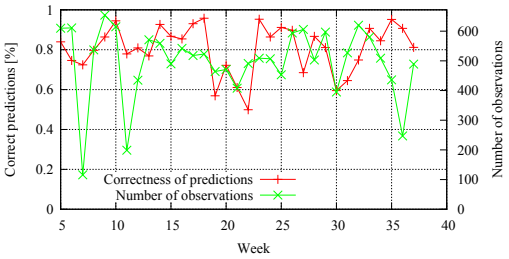
(b) Prediction errors for participant 120.



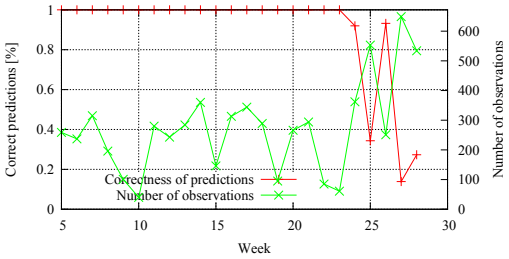
(c) Prediction errors for participant 123.



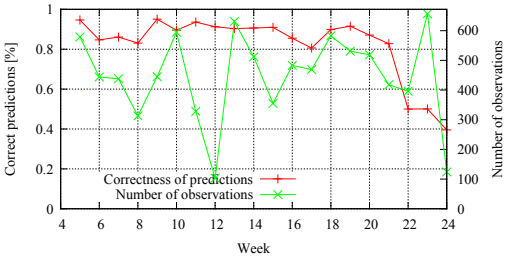
(d) Prediction errors for participant 126.



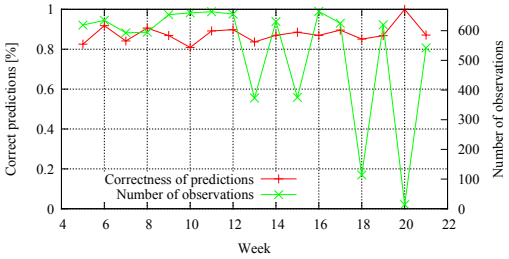
(e) Prediction errors for participant 127.



(f) Prediction errors for participant 139.

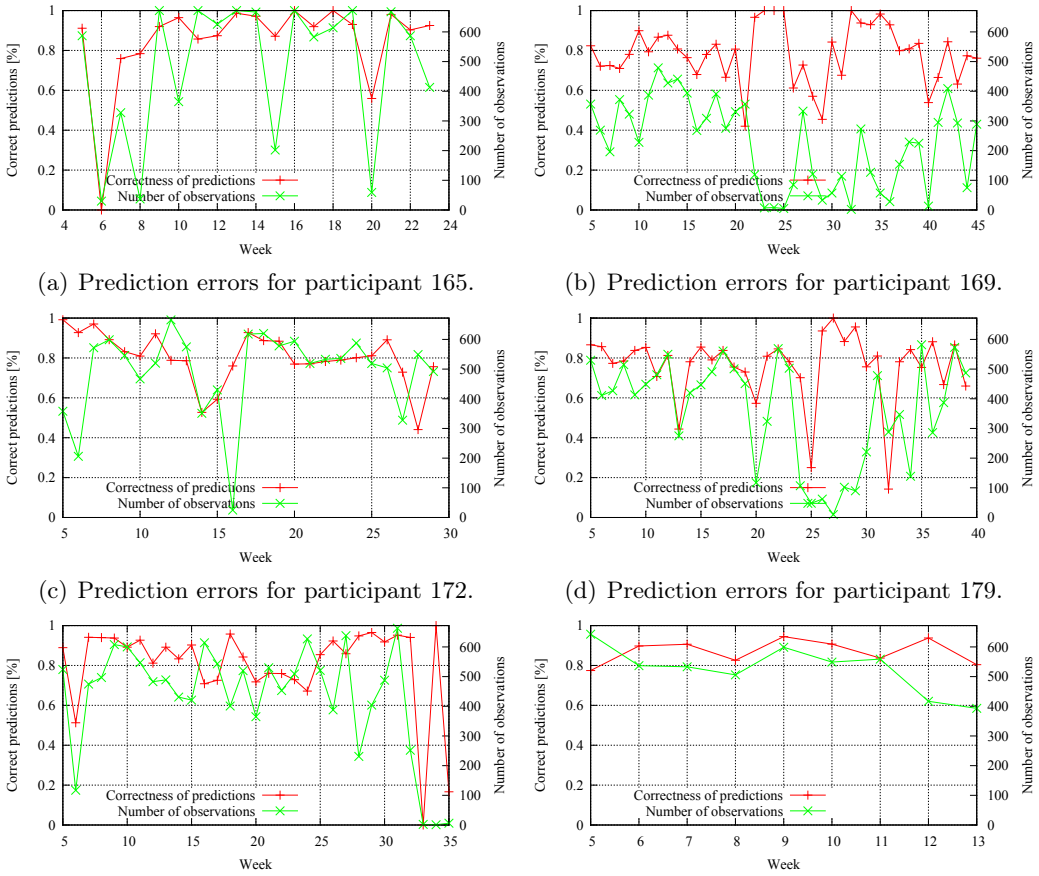


(g) Prediction errors for participant 141.



(h) Prediction errors for participant 160.

Figure 27: Prediction errors and number of observations for participants 117 to 160.



(a) Prediction errors for participant 165.

(b) Prediction errors for participant 169.

(c) Prediction errors for participant 172.

(d) Prediction errors for participant 179.

(e) Prediction errors for participant 185.

(f) Prediction errors for participant 186.

Figure 28: Prediction errors and number of observations for participants 165 to 186.