# HABITUAL BEHAVIOURAL PATTERN EXTRACTION BASED ON PASSIVE MOBILE TRACKING DATA

Sakura ARAKI [a], Fumitaka KURAUCHI [a], Toshiyuki NAKAMURA [a], Masahiro KUWAHARA [b]
[a] *Department of Civil Engineering, Gifu University, Japan*
[b] *Social system PF development div. InfoTech-AS, Toyota Motor Corporation, Tokyo, Japan*
*araki.sakura.c1@s.gifu-u.ac.jp*
*kurauchi.fumitaka.c3@f.gifu-u.ac.jp, ORCiD: 0000-0002-7448-8737*
*nakamura.toshiyuki.b9@f.gifu-u.ac.jp, ORCiD: 0000-0003-3093-4229*
*kuwahara@toyota-tokyo.tech, ORCiD: 0000-0003-2896-8622*

## 1. INTRODUCTION

Since the 1970s, one-day travel behaviour surveys have been conducted in Japan's major metropolitan areas, contributing to urban transportation policies. However, these ten-year interval, questionnaire-based surveys capture only a single day's behaviour from a small percentage of individuals, failing to account for variability and periodicity in daily activities. While past studies on travel behaviour analysis have shown that there is a certain repetitiveness or periodicity in daily activities, few have quantitatively analysed habitual patterns over longer periods, such as weeks or months. In addition, when looking at data collection methods, many of them require the active participation of individuals, such as questionnaires, activity diaries[1)] and user-input mobile apps[2)], which tend to have relatively small sample sizes. To address this, passive data collection methods are gaining attention, leveraging advancements in mobile communication and GPS technology to gather large-scale movement data. This study proposes a method to analyse habitual travel patterns using passively collected mobile tracking data. By identifying behaviours within weekly or monthly cycles, the approach captures fixed patterns amid daily fluctuations, enabling more accurate demand forecasting and optimised transportation services, such as revised bus schedules or new mobility routes tailored to regional demand needs.

## 2. METHODOLOGY

### 2.1 Data

We use "point-based population flow data" provided by Agoop Corporation. This big data, anonymised and processed from GPS data collected every few minutes via multiple smartphone apps, allows us to track people's movements as points. The study period is three months, from Thursday, September 1, 2022, to Thursday, November 30, 2022, focusing on Gifu City, Gifu Prefecture. The logs of 162,084 users with unique identification numbers for whole period who entered, exited, or stayed in the area at least once within Gifu City are extracted with information such as date, time, latitude, longitude, OS, gender, and age. To ensure data reliability, cleansing for GPS accuracy (≤40m) and speed (≤120 km/h). As a result, 238,466,085 records out of 280,046,212 records are found to be reliable point data.

## 2.2 Main Data Processing Step

The data consists of point sequences with movement and stay points. To estimate stay locations, we used the Python package "Trackintel"[3)] with three thresholds: (1) time threshold ≥ 60 minutes, (2) distance threshold ≥ 100 meters, and (3) time gap threshold (maximum time between two observations to be considered the same activity) ≤ 60 minutes. We focused on users who had staypoints for at least five days (22,057 users), resulting in 447,015 staypoints.

Staypoints were categorised into home, workplace, and other types using 500m mesh units. Home locations were determined using the first or last log of the day, calculating the centre of gravity, and identifying the corresponding 500m mesh. Workplace locations were estimated based on logs between 9:00–18:00, selecting the most frequently observed 500m mesh. For trip purposes, specific purposes (e.g., hospital visit, shopping) were not estimated. Instead, we used 500m mesh classifications from Watanabe and Kurauchi (2023)[4)], which categorised meshes in Gifu Prefecture into 19 types based on population and socioeconomic data. In this study, staypoints were classified by mesh characteristics as stays with similar characteristics, such as demographic composition and facility location. Therefore, staypoints were labelled as home (H), workplace (W), or one of mesh characteristics (19 cluster types).

## 2.3 Travel Pattern Extraction Method

To identify day-to-day travel patterns, we assume that people's behaviours are represented by regular patterns (weekly, bi-weekly, quad-weekly) and use the least squares method to identify each traveller's pattern. For each non-home stay cluster, a sequence of 84 binary variables is created, where a value of 1 indicates a visit on each day across 12 weeks (Sep 5 to Nov 27). As a result, a total of 56,578 combinations of users and staying clusters were obtained. Next, predefined cyclic patterns as {0,1} sequences are prepared. The patterns consist of seven weekly visitation patterns for each day of the week, 14 patterns for bi-weekly visits (considering starts in either the first or second week), 28 patterns for quad-weekly visits (considering starts from the first to the fourth week), and two patterns for visiting or not visiting on holidays, resulting in a total of 51 cyclic patterns.

Here let us define a sequence of visitation patterns with 84 components for staying cluster $c$ of individual $n$ as $\mathbf{x}_{cn}$. Also define a $i$th cyclic pattern as $\mathbf{p}_i$. For example, $\mathbf{p}_i$ for 'weekly Monday' pattern will be represented as $\mathbf{p}_i = (1,0,0,0,0,0,0,1,0,\dots)^T$ with values of 1 for 1st, 8th, 15th …,78th components. Also define $\mathbf{P}$ as a matrix with a size of (84 x 51) representing a concatenation of $\mathbf{p}_i$ as $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2 \dots, \mathbf{p}_{51})$. Define unknown binary variable vector for staying cluster $c$ of individual $n$, $\mathbf{y}_{cn}$, with the size of 51, whose components take a value of 1 the corresponding cyclic patterns are included in the observed data. We apply a least square method to minimise the difference between observed sequences and the overlay of cyclic patterns. The optimisation problem is defined as follows;

$$\min(\mathbf{x}_{cn} - \mathbf{P}\mathbf{y}_{cn})^T(\mathbf{x}_{cn} - \mathbf{P}\mathbf{y}_{cn}) \text{ , subject to } \mathbf{y}_{cn} = \{0,1\}.$$

## 3. Results

As a result of pattern extraction, approximately 20% (12,138 cases) of the 56,578 user-location combinations displayed a periodic pattern. In contrast, the remaining 44,440 instances were classified as having "No Patterns". Next, the visits that match the periodic patterns for each day were counted as "periodic visits," while visits that do not match the patterns and "no pattern" visits are counted as "non-periodic visits." A portion of the daily trends is shown (**Figure 1**), and the aggregated results by each day of the week are presented (**Table 1**). From these results, on weekdays, over 25% of visits have periodicity, while on weekends, the periodicity is lower than on weekdays. Specifically, Saturday has a periodic visit ratio of 20.3%, while Sunday is even lower at 18.3%. Additionally, Friday, with the highest number of 3,408 visits, shows the lowest periodic visit ratio among weekdays at 25.8%. Previous studies have similarly shown that weekday movements are more stable than those on weekends, that the difference in movement patterns between weekdays and Sundays is greater than that between weekdays and Saturdays[1),2)]. It has also been stated that Friday plays a role in broadening activity locations[5)]. Our findings align with these insights.

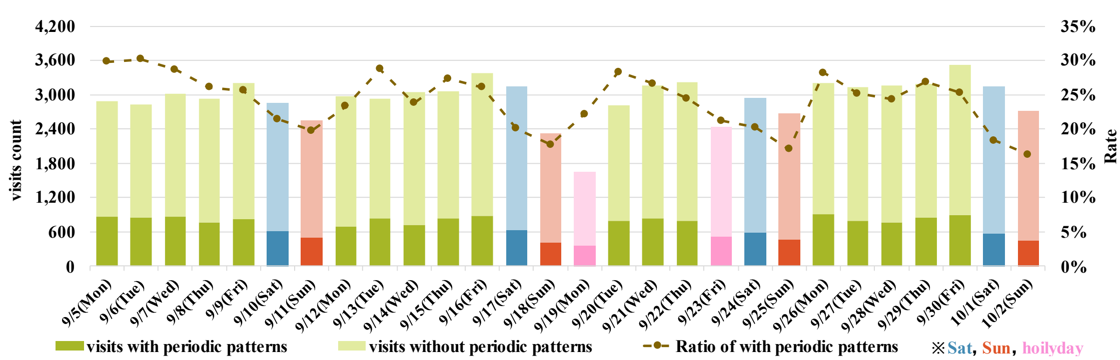


**Figure 1.** ***Visits with and without periodic patterns*** ***(Sep 5 to Oct 2)***

**Table 1.** ***Visits with and without periodic patterns per day and Ratio of with periodic patterns*** ***(Sep 5 to Nov 27)***

| | *Mon* | *Tue* | *Wed* | *Thu* | *Fri* | *Sat* | *Sun* | *holiday* |
|---|---|---|---|---|---|---|---|---|
| *Visits count /day* | *3,118* | *3,123* | *3,198* | *3,190* | *3,408* | *3,094* | *2,611* | *2,351* |
| *Visits with periodic patterns /day* | *836* | *876* | *844* | *848* | *880* | *629* | *478* | *462* |
| *Visits without periodic patterns /day* | *2,281* | *2,248* | *2,353* | *2,342* | *2,528* | *2,465* | *2,133* | *1,889* |
| ***Ratio of with periodic patterns*** | ***26.8%*** | ***28.0%*** | ***26.4%*** | ***26.6%*** | ***25.8%*** | ***20.3%*** | ***18.3%*** | ***19.9%*** |

To clarify which periodic patterns are most prevalent, we calculated the pattern shares for visits on each day of the week, excluding holidays (**Table 2**). As a result, quad-weekly visits were the most frequent, making up about 14%. Weekly and bi-weekly visits varied between weekdays and weekends: on weekdays, Weekly visits accounted for around 10%, while bi-weekly visits accounted for about 3%. Focusing on weekdays, Monday-Tuesday and Wednesday-Thursday-Friday groups had similar distributions, but the latter group showed a slight decrease in Weekly visits and an increase in bi-weekly visits.

**Table 2.** ***Ratios with periodic patterns for each day of the week*** ***(Sep 5 to Nov 27)***

| | | Mon | Tue | Wed | Thu | Fri | Sat | Sun |
|---|---|---|---|---|---|---|---|---|
| ***Ratio of with periodic patterns*** | | ***26.8%*** | ***28.0%*** | ***26.4%*** | ***26.6%*** | ***25.8%*** | ***20.3%*** | ***18.3%*** |
| | ***Weekly*** | *10.9%* | *11.1%* | *9.1%* | *9.3%* | *8.3%* | *4.2%* | *3.2%* |
| | ***Bi-weekly*** | *2.1%* | *2.8%* | *3.7%* | *3.5%* | *3.7%* | *1.8%* | *1.2%* |
| | ***Quad-weekly*** | *13.8%* | *14.1%* | *13.6%* | *13.8%* | *13.8%* | *14.3%* | *13.9%* |

## 4. Conclusion

This study explored a method for understanding users' habitual movement patterns using mobile tracking data passively collected from mobile devices. Specifically, it focused on periodic visits related to specific weekdays, including weekly, bi-weekly, and quad-weekly cycles. However, methods for quantitatively evaluating habitual behaviour patterns are still not fully established, and there is currently limited knowledge to support the results of this analysis. Nevertheless, this study provides a framework for quantitatively assessing the extent of periodic movement within the fluctuating daily movement demand using a statistical and objective approach. This approach enables a deeper understanding of multi-day periodicity, which traditional movement behaviour surveys were unable to capture, offering a new perspective in movement behaviour analysis. Future research involves comparisons of analysis results using other location data, improvements in spatiotemporal resolution, and simulations of demand estimation that incorporate periodicity as a feature.

## References

1) Schlich, R. and Axhausen, K.W. Habitual travel behaviour: Evidence from a six-week travel diary. Transportation, 30, 13–36 (2003).
2) Xianyu, J., Rasouli, S. and Timmermans, H. Analysis of variability in multi-day GPS imputed activity-travel diaries using multi-dimensional sequence alignment and panel effects regression models. Transportation, 44(3), pp. 533–553 (2017).
3) Bucher, D., Martin, H., Hong, Y., and Wiedemann, N. 2019. The trackintel Documentation – trackintel 1.3.1 documentation. URL：https://trackintel.readthedocs.io/en/latest/
4) Watanabe, A. and Kurauchi, F. Characterization of the region in Gifu Prefecture using mobile spatial statistical data, Infrastructure Planning Review 68, 7046 (2023).
5) Susilo, Y.O. and Kitamura, R. Analysis of Day-to-Day Variability in an Individual's Action Space: Exploration of Six-Week Mobidrive Travel Diary Data. Transportation Research Record Journal of the Transportation Research Board 1902(1) (2005).