Data-Driven Discovery of Mobility Periodicity for Understanding Urban Systems

Xinyu Chen
Massachusetts Institute of Technology

Qi Wang
Northeastern University

Yunhan Zheng
Massachusetts Institute of Technology

Nina Cao
Massachusetts Institute of Technology

HanQin Cai
University of Central Florida

Jinhua Zhao*
Massachusetts Institute of Technology

September 1, 2025

ABSTRACT

Human mobility regularity is crucial for understanding urban dynamics and informing decision-making processes. This study formulates the periodicity quantification problem in complex human mobility data as a sparse identification of dominant positive auto-correlations in time series autoregression, allowing one to quantify significant periodic patterns. We apply the framework to real-world metro passenger flow in Hangzhou, China and multi-modal mobility data in New York City and Chicago, USA, revealing the interpretable weekly periodicity across different spatial locations over past several years. The analysis of ridesharing data from 2019 to 2024 demonstrates the disruptive impact of the pandemic on mobility regularity and the subsequent recovery trends. This study also uncovers mobility patterns of different travel modes such as ridesharing, taxi, subway, and bikesharing in Manhattan. Our findings highlight the potential of interpretable machine learning to discover spatiotemporal mobility patterns and offer a valuable tool for understanding urban systems.

Keywords Human mobility · Transportation systems · Regularity · Periodicity · Time series · Machine learning · Optimization

1 Introduction

Human mobility in cities often reveals recurring patterns—the predictable ebbs and flows of people moving through urban areas—that reflect the underlying structure of urban life [Gonzalez et al., 2008, Song et al., 2010, Simini et al., 2012]. From daily commutes to weekend trips, these temporal regularities in human mobility underpin everything from urban transportation management [Sheffi, 1985] and transit scheduling [Guihaire and Hao, 2008] to predicting the spread of diseases [Balcan et al., 2009, Belik et al., 2011, Castillo-Chavez et al., 2016, Li et al., 2022, Du et al., 2018] and optimizing resource allocation [Katoh et al., 2024]. In essence, human mobility regularity is an important concept for identifying periodic trips in urban areas, which often follow daily and weekly cycles. A clear example is daily commuting, where passenger flows typically exhibit morning and afternoon peaks, forming a roughly "M"-shaped pattern in time series data [Chen and Sun, 2022] (see e.g., Fig. 1B). As cities evolve, mobility patterns have become increasingly complex and less predictable, characterized by complicated regulatory policies, diverse transportation modes and behavioral shifts driven by social economic and environmental factors [Midgley, 2009, Machado et al., 2018, Tirachini, 2020, Nouvellet et al., 2021, Li et al., 2022]. Thus, understanding the rhythm of human mobility in urban areas is essential for grasping how our societies function and urban systems operate [Gonzalez et al., 2008, Song et al., 2010, Simini et al., 2012, Batty et al., 2012, Jiang et al., 2017, Alessandretti et al., 2020].

A paradigm for discovering long-term human mobility relies on collecting and constructing time series with a certain time resolution (e.g., hourly resolution) across different spatial locations over different years, which are often characterized as an algebraic tensor structure in Fig. 1A. In practice, tensors are multidimensional arrays which have been widely used to represent time series data of human mobility [Chen and Sun, 2022]. To explore human mobility regularity underlying these time series, a few critical but unsolved questions arise as follows. First, there is no standard method to quantify and compare the strength of mobility regularity and periodicity with different noise levels and temporal structures. The two example time series of passenger flow in Figs. 1B-C show different periodic patterns, while the
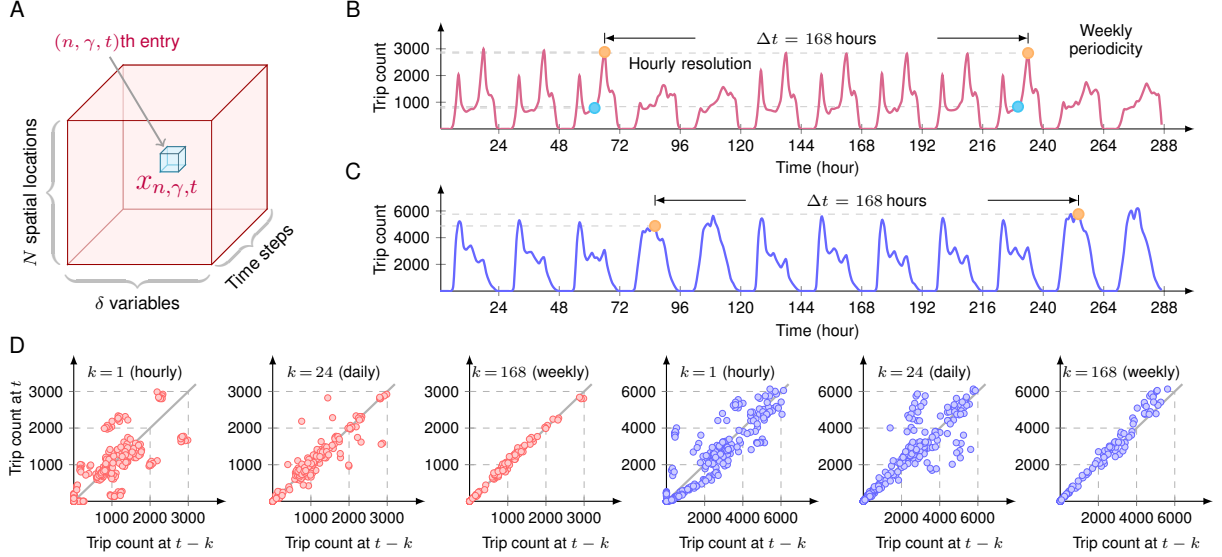
---

*Corresponding author. Email: `jinhua@mit.edu`

Figure 1: Illustration of time series autoregression and its multidimensional setting on human mobility data. (A) The dth-order autoregression on the univariate time series can be formulated as an optimization problem for finding the optimal coefficients. Here, the order d is an integer, determining the number of auto-correlations. (B) Multidimensional time series autoregression. Note that the symbol "$\times$" simply refers to the element-wise product between matrices. (C) The interpretability of time series autoregression can be reinforced by sparse and non-negative coefficients. The upper bound of the number of nonzero entries in the coefficient vector is set as an integer τ, which is formally defined as the sparsity level. In mathematics, the cardinality of index set Ω (i.e., indices of nonzero coefficients) can be used to reformulate sparsity constraints. As a result, the decision variables in the constructed optimization problem include both coefficients and index set.

strength of periodicity can be intuitively revealed by the auto-correlation analysis in Fig. 1D. Although it is obvious to observe from Fig. 1D that the weekly periodicity of Fig. 1B is stronger than Fig. 1C, the question that remains is how to compute the comparable metric for time series periodicity by referring to the auto-correlation plots in Fig. 1D. Second, there is no established approach to leverage periodicity strength as a metric for tracking the spatial and temporal evolution of mobility systems. Previous research has adopted methods such as Fourier Transform [Li et al., 2010, Prabhala et al., 2014], entropy-based measures [Goulet-Langlois et al., 2017, Huang et al., 2019, do Couto Teixeira et al., 2021, Teixeira et al., 2021], reinforcement learning [Tao et al., 2021], and clustering of similar patterns [Manley et al., 2018, Zhong et al., 2015, Zhang and Song, 2022] to identify periodicity and regularity in human mobility. In dynamical systems, dynamic mode decomposition identifies the cyclical patterns through the eigenvalue decomposition of vector autoregressive coefficients [Brunton and Kutz, 2022].

However, these approaches typically do not provide an explicit and interpretable metric that directly quantifies periodicity, making them less suitable for downstream modeling or individualized interpretation. In addition, most studies that quantify periodicity or regularity in human mobility data focus on the analysis of individual trajectory data rather than aggregated demand or travel-mode usage data [Prabhala et al., 2014, Goulet-Langlois et al., 2017, Huang et al., 2019, Teixeira et al., 2021, do Couto Teixeira et al., 2021, Tao et al., 2021, Manley et al., 2018, Oliveira et al., 2016]. While individual data can provide useful information on predicting mobility flows, interpretable information on aggregated demand data reveals longer term trends that can be more directly related to larger economic, social, and other changes that can help inform the development of policies in the future. In urban systems, human mobility patterns vary with different travel modes such as public transit and ridesharing. It is also meaningful to fairly quantify mobility regularity of different travel modes and utilize periodicity patterns to allocate resources and inform urban planning. Another fact is that mobility flows can be severely disrupted by extreme events and may evolve gradually over time due to changes in behavior, infrastructure, or policy. For example, the COVID-19 pandemic disrupted regular travel routines and led to a noticeable decline in mobility regularity [Li et al., 2022, Chen et al., 2025a]. The challenge is how to systematically use periodicity strength to detect, quantify, and interpret such changes across both space and time.

Substantial progress has been made in understanding human mobility regularity and patterns from analytical perspectives such as entropy and predictability [Oliveira et al., 2016, Barbosa et al., 2018, Delussu et al., 2023]. However, the

existing methods are not well-suited to quantifying periodicity across spatiotemporal domains and discovering dynamic patterns that related to human mobility regularity. Recently, interpretable machine learning methods [Murdoch et al., 2019, Rudin et al., 2022, Brunton and Kutz, 2022] offer new opportunities to improve appearance, robustness, and interpretability of data-driven exploration in mobility data. In terms of model development, sparse linear regression [Tibshirani, 1996, Jenatton et al., 2011] can be used to reformulate the problem of sparse identification of nonlinear dynamical systems [Brunton et al., 2016], which laid the foundations of several machine learning frameworks that designed for variable and feature selection in scientific discovery. In that venue, sparsity is one of the most important aspects to reinforce interpretability of machine learning models [Murdoch et al., 2019, Rudin et al., 2022]. However, existing sparse methods face critical limitations when applied to large-scale human mobility data. First, classical sparse regression methods are primarily designed for variable selection from univariate and multivariate data and struggle to scale to multidimensional data with spatial and temporal dependencies. Second, mobility data are inherently noisy, irregular, and heterogeneous across spatiotemporal domains, which challenges standard sparse models that lack built-in mechanisms for structured robustness. Third, even when sparsity is enforced, the resulting models do not yield interpretable or comparable metrics for periodicity strength across different spatial locations, variables, and time phases.

To bridge these gaps, we propose a multidimensional sparse autoregression framework that models time series of mobility flows across space and time, explicitly captures dominant auto-correlations, and provides a unified and interpretable measure of periodicity strength. This interpretable machine learning framework builds upon the tensor data structure. Formally, human mobility data are spatially- and temporally-varying, and can be represented as a third-order tensor of non-negative integers $x_{n,\gamma,t} \in \{0, 1, 2, \cdots\}$ as shown in Fig. 1A, where $n \in [N]$ indexes spatial locations, $\gamma \in [\delta]$ indexes different variables or time phases, and $t \in [T_\gamma]$ indexes time steps. This algebraic tensor structure captures the multidimensional nature of mobility systems and aligns with a rich body of work on tensor-based machine learning for spatiotemporal pattern discovery [Kolda and Bader, 2009, Chen et al., 2024]. Each individual sequence $x_{n,\gamma,t}$ is a time series, and thus can be seamlessly characterized by using time series models such as autoregression [Hamilton, 2020]. As exemplified by Fig. 1B, real-world passenger flow time series demonstrate weekly periodicity by picking data points with a weekly cycle. When these data points are denoted by $x_t, t \in [T]$, the auto-correlations in Fig. 1D can be formulated as $x_t \approx w_k x_{t-k}$, measuring self-similarity in mobility over any time lag k. Observing Fig. 1D, data points of the example time series are very close to the anti-diagonal curve at a weekly cycle (i.e., $k = 168$), implying that weekly periodicity is positively auto-correlated. Moreover, the "closeness" of data points to the anti-diagonal at index $k = 168$ in Fig. 1D justifies the strength of weekly periodicity of time series, demonstrating that the time series in Fig. 1C is less periodic at a weekly cycle than Fig. 1B.

2 Machine Learning Formula

Interpretable machine learning provides a way to understand underlying patterns of complicated data and systems [Brunton et al., 2016, Murdoch et al., 2019]. This study starts from the core idea of autoregression on time series, namely, the current value of the time series can be predicted as a linear combination of its past values [Hamilton, 2020]. The objective is to find a set of coefficients that minimizes the sum of squared autoregressive errors (Fig. 2A) with predefined constraints such as sparsity and non-negativity on the coefficients as shown in Fig. 2C. Sparsity, enforced by the constraint $|\Omega| \leq \tau$, limits the number of nonzero coefficients to a small integer τ. This implies that only a few past time lags significantly contribute to predicting the current value. For instance, Fig. 1D shows that the trip counts at time t are highly consistent with the trip counts at time $t - 168$, revealing predictable and periodic patterns.

In the context of human mobility periodicity, we expect dominant nonzero coefficients that uncover periodic patterns such as a 168-hour cycle for weekly periodicity. Non-negativity of the coefficients ensures that the identified auto-correlations are positive, aligning with an intuitive understanding of periodicity as similar patterns repeating over time (see e.g., Fig. 1D). Multidimensional time series autoregression follows the specification of auto-correlations in which the coefficients over N spatial locations and δ variables can be optimized (Fig. 2B). The optimization explicitly aims to find a small set of relevant past time lags in the index set Ω over $N \times \delta$ coefficient vectors. Formally, on the human mobility data $\{x_{n,\gamma,t}\}_{n\in[N],\gamma\in[\delta],t\in[T_\gamma]}$, we formulate the multidimensional sparse autoregression as follows,

$$\min_{\{w_{n,\gamma,k}\},\Omega} \sum_{n=1}^{N} \sum_{\gamma=1}^{\delta} \sum_{t=d+1}^{T_\gamma} \left(x_{n,\gamma,t} - \sum_{k\in\Omega} w_{n,\gamma,k} x_{n,\gamma,t-k} \right)^2$$

$$\text{subject to} \begin{cases} |\Omega| \leq \tau, \ \tau \in \mathbb{Z}^+, \\ w_{n,\gamma,k} \geq 0, \forall k \in \Omega, \\ w_{n,\gamma,k} = 0, \forall k \notin \Omega, \\ \sum_{k\in\Omega} w_{n,\gamma,k} = 1, \end{cases} \quad (1)$$
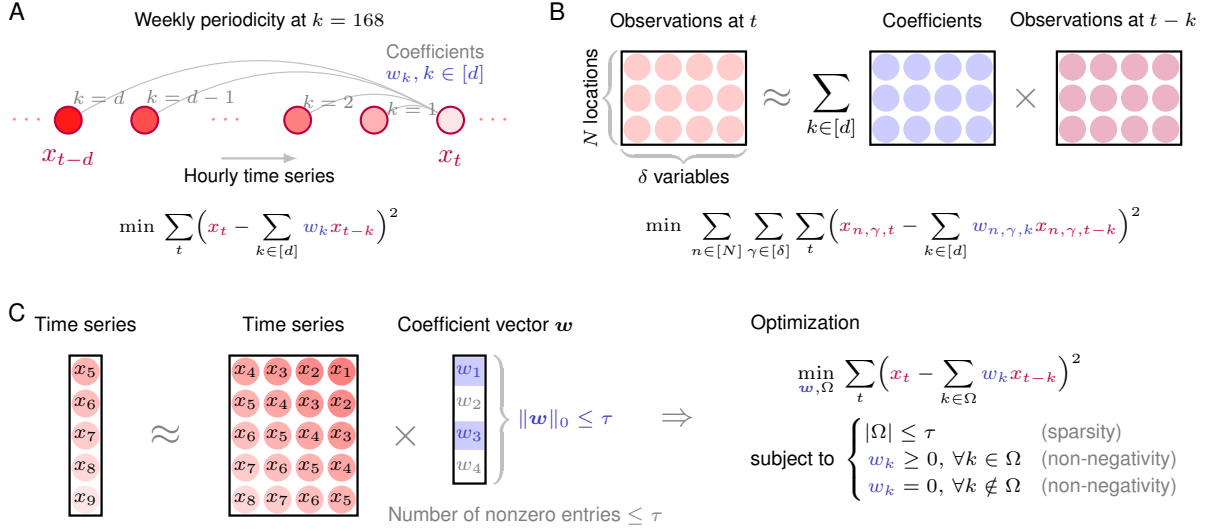
3

Figure 2: Illustration of time series autoregression and its multidimensional setting on human mobility data. (A) The dth-order autoregression on the univariate time series can be formulated as an optimization problem for finding the optimal coefficients. Here, the order d is an integer, determining the number of auto-correlations. (B) Multidimensional time series autoregression. Note that the symbol "$\times$" simply refers to the element-wise product between matrices. (C) The interpretability of time series autoregression can be reinforced by sparse and non-negative coefficients. The upper bound of the number of nonzero entries in the coefficient vector is set as an integer τ, which is formally defined as the sparsity level. In mathematics, the cardinality of index set Ω (i.e., indices of nonzero coefficients) can be used to reformulate sparsity constraints. As a result, the decision variables in the constructed optimization problem include both coefficients and index set.

where $w_{n,\gamma,k}, n \in [N], \gamma \in [\delta], k \in [d]$ are the non-negative coefficients of this dth-order autoregression formula, allowing one to quantify the positive auto-correlations such as daily and weekly periodicity as the dominant ones in time series. The last constraint is the normalization on the non-negative coefficients, letting the sum of the coefficients be one. Since the optimization problem is equivalent to the ℓ_0-norm induced sparse autoregression, it can be addressed by mixed-integer optimization solvers with exact solutions [Bertsimas et al., 2016, 2020, 2024, Tillmann et al., 2024]. On such multidimensional tensor time series, we can discover the weekly periodicity of human mobility across different spatial locations and variables (or time phases).

On the hourly time series, if one intends to identify weekly periodicity, then the order of sparse autoregression should be set as $d \geq 168$ or greater. In what follows, we consider the order $d = 168$ for the real-world mobility datasets. While the constraint $|\Omega| \leq \tau$ requires selecting a subset of τ coefficients out of d possible ones, we extract the coefficients at index $k = 168$, namely, auto-correlations $w_{n,\gamma,168}, n \in [N], \gamma \in [\delta]$ to interpret the strengths of weekly periodicity across N spatial locations over δ variables. The sparsity level τ is a critical hyperparameter for sparse autoregression. If this value is very small, then the index $k = 168$ might not be in the support set Ω, making it difficult to quantify weekly periodicity. In contrast, given a large sparsity level τ, the values of auto-correlations within the support set Ω would have marginal differences.

3 Metro Passenger Flow Periodicity

To demonstrate the effectiveness of our framework in capturing periodicity from real-world mobility data, we begin with a case study of metro passenger flow data in Hangzhou, China. This dataset offers a clean and well-structured setting with strong known periodic patterns, making it an ideal testbed for validating the ability of our sparse autoregression model to recover meaningful temporal cycles—particularly weekly periodicity across different stations and flow directions.

The metro passenger flow dataset in Hangzhou[2] was processed as a tensor structure with 81 stations, 2 variables (i.e., inflow and outflow), and 576 hours (i.e., over 24 days). Weekly periodicity is quantified by the non-negative

---

[2]The anonymized trip records are available at `https://doi.org/10.5281/zenodo.3145404`.
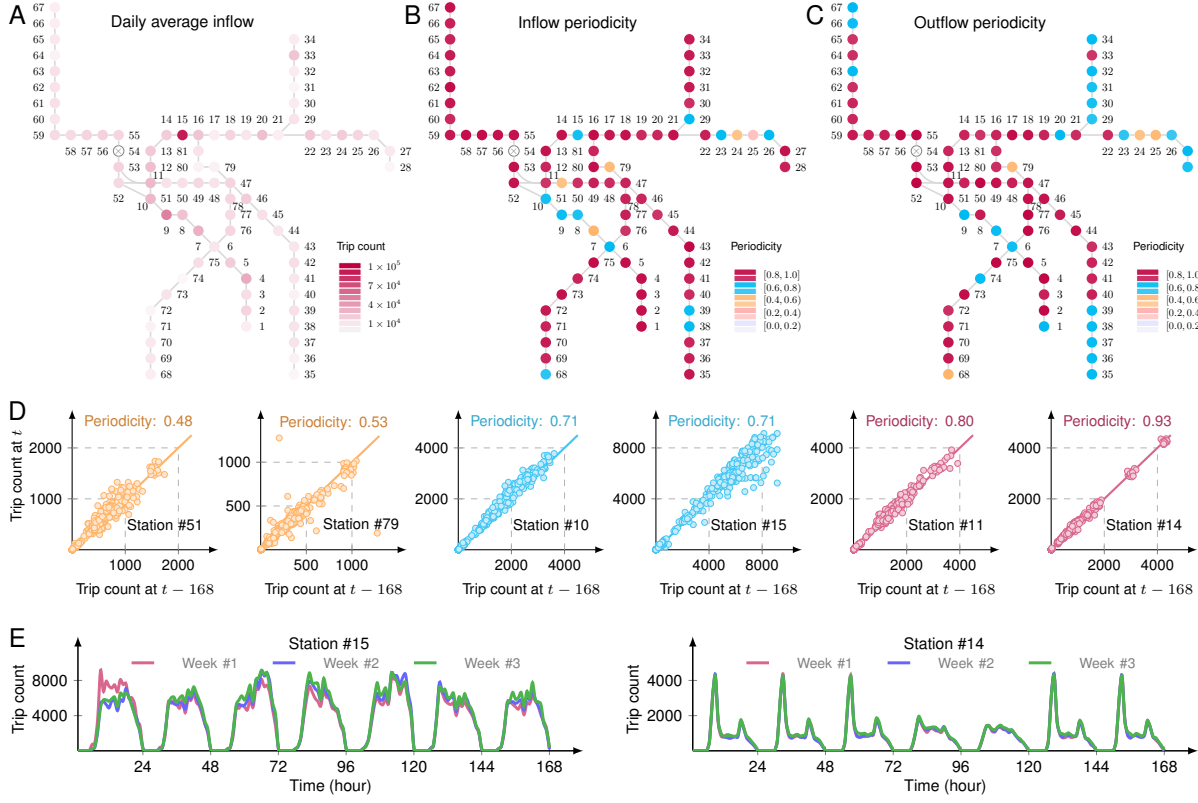
Figure 3: Identification of weekly periodicity in inflow and outflow time series from the Hangzhou metro passenger flow dataset (from January 2 to 25, 2019) with the multidimensional sparse autoregression method (sparsity level $\tau = 4$). (A) Daily average inflow trips of 81 metro stations. Note that the trip records of the station #54 are missing. (B) Weekly periodicity of hourly inflow time series achieved by the sparse autoregression with sparsity level $\tau = 4$. 64 stations show high periodicity (≥ 0.8). The average periodicity value is 0.844. (C) Weekly periodicity of hourly outflow time series. 53 stations show high periodicity (≥ 0.8). The average periodicity value is 0.829. Since inflow and outflow are two variables in the tensor data, their coefficients are optimized simultaneously with the same index set. (D) Scatter plots of trip counts at time t versus time $t - 168$ of 6 selected stations, illustrating varying strengths of weekly periodicity. The time series with a high weekly periodicity value also presents a strong similarity of data points between times t and $t - 168$. (E) Hourly inflow time series for stations #14 and #15, shown across three consecutive weeks: January 2–8, January 9–15, and January 16–22. Strong overlap among these weekly segments indicates high periodicity at a weekly cycle.

coefficients at index $k = 168$ in the multidimensional sparse autoregression with the prescribed order $d = 168$. In the experiments, we set the sparsity level as $\tau = 4$ depending on how many dominant auto-correlations we expect to identify. In the mixed-integer optimization solver, the support set Ω was optimized simultaneously with the coefficients. In this case, the support set is optimized as $\Omega = \{1, 24, 144, 168\}$, corresponding to hourly, daily, 6-day, and weekly cycles, respectively. The coefficients at index $k = 168$ refer to as the strength of weekly periodicity. The analysis of weekly periodicity of metro passenger flow reveals the mobility patterns of both inflow and outflow variables. Table 1 summarizes the detailed comparison of weekly periodicity between inflow and outflow variables. Observing Figs. 3B-C and Table 1, the inflow time series at the end stations of metro lines (e.g., $\{27, 28\}$, $\{30, 31, 32, 34\}$, $\{35, 36, 37\}$, and $\{63, 66, 67\}$) showed stronger periodicity than the outflow time series. Conversely, the outflow time series of stations in downtown areas (e.g., $\{7, 8, 10, 15, 51\}$) were found to be more periodic than the inflow time series. These findings are relevant to commuting trips in urban areas.

Fig. 3D shows the weekly periodicity values for six selected stations, measured as the strength of auto-correlations at a weekly lag (i.e., $k = 168$) in the inflow time series. The source emphasizes that time series with high weekly periodicity values also show a strong similarity between data points at times t and $t - 168$, indicating positive auto-correlations. The weekly periodicity is visually confirmed by scatter plots and time series overlaps. For instance, station #14 shows

Table 1: Strength of weekly periodicity of the Hangzhou metro passenger flow data at the selected end stations of metro lines.

Station #	Inflow	Outflow	Station #	Inflow	Outflow
23	**0.77**	0.65	35	**0.91**	0.76
24	0.46	**0.48**	36	**0.88**	0.74
25	0.24	**0.54**	37	**0.85**	0.74
26	**0.78**	0.62	38	**0.76**	0.75
27	**0.93**	0.78	39	0.73	**0.74**
28	**0.87**	0.69	40	0.83	**0.88**
29	**0.77**	0.66	62	**0.95**	0.82
30	**0.81**	0.71	63	**0.89**	0.77
31	**0.90**	0.70	64	**0.89**	0.81
32	**0.91**	0.74	65	**0.86**	0.80
33	**0.90**	0.80	66	**0.87**	0.73
34	**0.90**	0.76	67	**0.93**	0.79

a high weekly periodicity of 0.93, and its scatter plot shows a strong alignment along the anti-diagonal (Fig. 3D). In particular, the comparison between stations #14 and #15 in Fig. 3E answers why the strength of time series periodicity is a fair metric to reveal repeating patterns of human mobility. The periodicity value of station #15 is 0.71, which is smaller than station #14. As shown in Fig. 3E, the trip count time series at station #15 does not align perfectly across weeks, indicating lower weekly periodicity compared to the more consistent pattern observed at station #14. To summarize, on the passenger flow dataset, the comparable metric of weekly periodicity across different stations and inflow/outflow variables to discover periodic mobility patterns in a city-wide scale.

4 Regularity of Ridesharing Trips

Multidimensional sparse autoregression quantifies human mobility periodicity through coefficients. The resulting non-negative coefficients directly quantify the strength of the positive auto-correlation at certain time lags. A high coefficient at a time lag corresponding to a weekly cycle (e.g., 168 hours for hourly time series data) indicates a strong weekly periodicity.

As we know, the COVID-19 pandemic had a significant disruptive impact on urban mobility regularity and periodicity. To examine how such disruptions affected periodicity, we analyze large-scale ridesharing trip data from 2019 to 2024 in NYC[3] and Chicago[4]. This analysis aims to demonstrate how the strength of weekly periodicity varied across different urban areas and years, reflecting the impact of the pandemic. In the experiments, we set the sparsity level and the order of sparse autoregression as $\tau = 6$ and $d = 168$, respectively. On both NYC and Chicago ridesharing datasets, the support set is optimized as $\Omega = \{1, 23, 24, 143, 167, 168\}$, covering the indices of local auto-correlations, daily cycle, and weekly cycle.

In NYC, the weekly periodicity of pickup trips in 2019 was similar to the patterns observed in 2022–2024, indicating a return to pre-pandemic regularity after disruptions (Fig. 4A). The year 2020 marked a significant shift from this regularity in NYC, with the pandemic severely disrupting trip patterns and reducing overall periodicity strength. By 2021, the ridesharing trips in NYC began recovering, transitioning from pandemic-induced irregularity towards the post-pandemic norm. During this recovery phase, suburban areas displayed increasingly periodic trip patterns in NYC. The periodicity patterns of ridesharing trips in 2022 became consistent with the patterns in 2019, while both periodicity values are almost same. Across these 6 years, ridesharing trips in downtown areas such as Manhattan were less periodic than suburban areas. In 2024, the overall weekly periodicity of NYC has been reinforced slightly. The ridesharing trips in suburban areas tend to be more periodic than 2019, while the downtown areas were less periodic (Fig. 4C).

In Chicago, a similar pattern emerged where the overall periodicity of ridesharing trips in 2019 was stronger than in 2020, and downtown areas in 2020 were less periodic than in 2019 (Fig. 4B). However, in contrast to NYC, the periodicity pattern transition in Chicago was slower, with no remarkable pattern change between 2020 and 2021. The periodicity patterns of ridesharing trips in Chicago during 2022–2024 also became consistent with patterns in 2019, suggesting a return to pre-pandemic regularity. In 2024, although the ridesharing trips demonstrated the same level of weekly periodicity as 2019, downtown areas became less periodic (Fig. 4C). Comparing the aggregated trip counts in

---

[3]The NYC ridesharing trip record dataset is available at `https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page`.

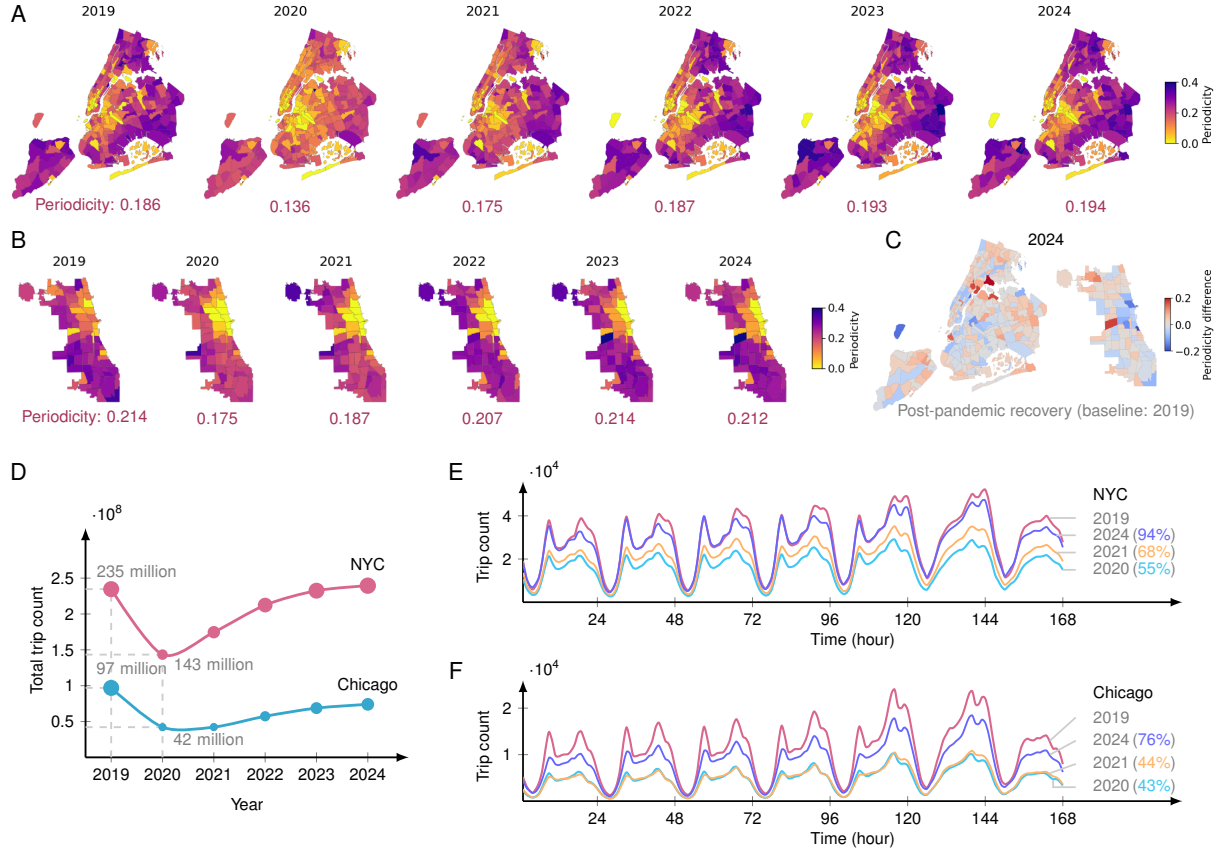[4]The Chicago ridesharing trip record dataset is available at `https://data.cityofchicago.org/`.

Figure 4: Identification of weekly periodicity in ridesharing trip data across different urban areas with the multidimensional sparse autoregression method (sparsity level $\tau = 6$). (A) Weekly periodicity of NYC ridesharing pickup trip data across 265 areas from 2019 to 2024. The average periodicity values are also marked. (B) Weekly periodicity of Chicago ridesharing pickup trip data across 77 areas from 2019 to 2024. (C) The coefficient difference between 2024 and 2019, referring to the coefficients of 2024 minus 2019. (D) Annual ridesharing trip counts of NYC and Chicago from 2019 to 2024. In contrast to NYC, the ridesharing trips in Chicago show a poor recovery in the post-pandemic years. Note that the ridesharing trip data of January 2019 is not available from the data portal. (E) Weekly average trip time series of NYC ridesharing, starting from Monday to Sunday. (F) Weekly average trip time series of Chicago ridesharing, starting from Monday to Sunday. While NYC shows the recovery percentage as 68% in 2021, Chicago keeps 44%, roughly as the same level of 2020.

Fig. 4D, NYC showed a quick return of ridesharing trip counts and reached the same level as 2019, while Chicago showed a poorer recovery in the post-pandemic years. The weekly average trip time series also indicated that while NYC showed a recovery percentage of 68% in 2021 (Fig. 4E), Chicago remained at 44%, similar to its 2020 level (Fig. 4F). Although Chicago showed a recovery percentage of 76% ridesharing trips in 2024 (Fig. 4F), the weekly periodicity already reached the same level as 2019.

Our analysis so far suggests that the COVID-19 pandemic led to a significant decrease in the periodicity of urban mobility in both NYC and Chicago in 2020. Both cities showed signs of recovery in subsequent years, with periodicity patterns returning to pre-pandemic levels by 2022-2024. However, the pace of recovery in terms of both periodicity and overall trip counts appeared to be faster in NYC compared to Chicago. Furthermore, our method enables policymakers to measure disruptions such as pandemics and design targeted recovery strategies.

5   Mobility Periodicity of Multiple Travel Modes

In Manhattan, multi-modal trip records from 2024 reveal distinct periodicity across ridesharing, yellow taxi, subway, and bikesharing (Fig. 5). This analysis aims to demonstrate the differences of weekly periodicity across four travel
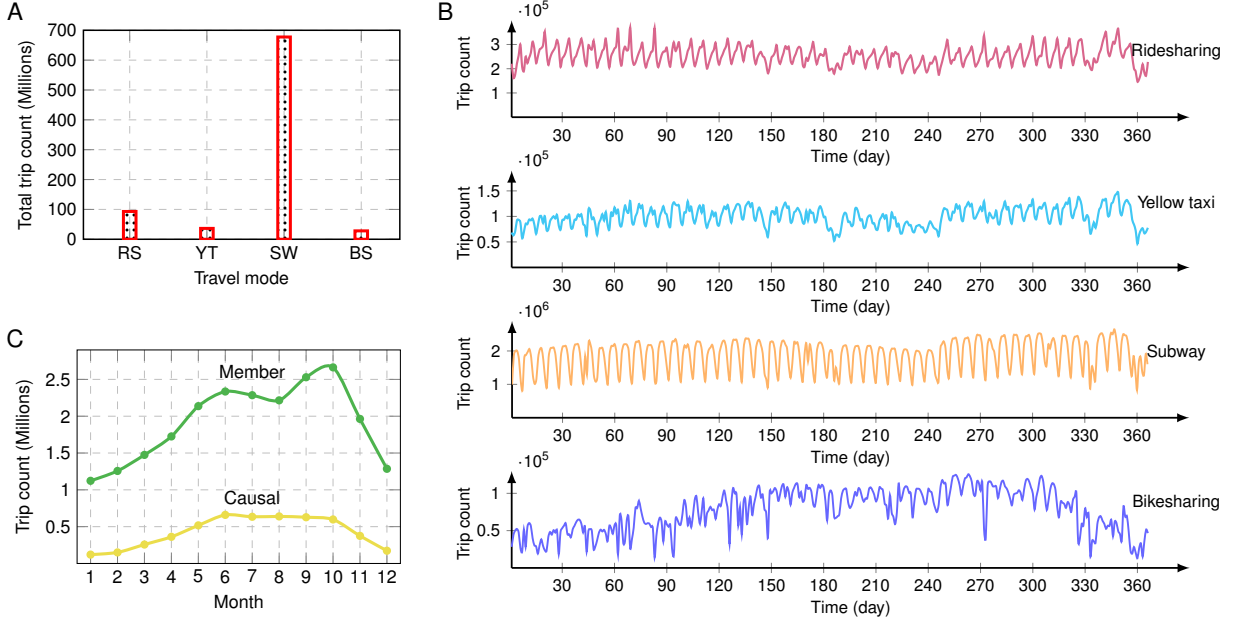
Figure 5: Trip counts of ridesharing (RS), yellow taxi (YT), subway (SW), and bikesharing (BS) in Manhattan of 2024. (A) The total trip counts of ridesharing, yellow taxi, subway, and bikesharing are 93.1 million, 36.4 million, 677.5 million, and 28.1 million, respectively. (B) Trip time series with a daily resolution in Manhattan. (C) Bikesharing trips of membership and causal users across 12 months of 2024.

modes and the spatiotemporal patterns of multi-modal mobility data. As shown in Fig. 5A, total trip volumes decrease in the order of subway (677.5 million), ridesharing (93.1 million), yellow taxi (36.4 million), and bikesharing (28.1 million). The daily time series in Fig. 5B indicates that subway ridership exhibits the most pronounced periodicity, with highly regular weekly cycles sustained throughout the year. In contrast, ridesharing, yellow taxi, and bikesharing display lower periodic structures and greater short-term variability. Fig. 5C further disaggregates bikesharing trips into membership and casual users. Membership trips constitute 23.0 million rides, or approximately 82% of the total, while casual trips account for the remaining 18%. A strong seasonal pattern is evident, with trip counts peaking during the warmer months (May–October) and declining sharply in winter (January, February, December), reflecting the sensitivity of cycling activity to weather conditions.

As shown in Fig. 6A–B, subway and bikesharing trip records are spatially projected onto 69 areas of Manhattan, yielding trip time series at the same spatial resolution as ridesharing and yellow taxi. Treating each travel mode as a variable, the hourly trip time series across spatial areas form a tensor (Fig. 1A). Given the distinct mobility patterns of bikesharing members and casual users, the variable dimension of mobility tensor comprises ridesharing, yellow taxi, subway, bikesharing (member), bikesharing (casual), and bikesharing (all). The resulting dataset is a tensor of size $69 \times 6 \times 8784$, with the temporal dimension spanning all 8,784 hours of 2024. By using the multidimensional sparse autoregression in Fig. 2, the support set is optimized as $\Omega = \{1, 23, 24, 144, 167, 168\}$ where $k = 168$ refers to as a weekly cycle. The auto-correlations at the index $k = 168$ allow one to quantify the weekly periodicity of trip time series across different spatial areas and travel modes. Fig. 6C shows that subway trips are more periodic than other travel modes. Although the trips of ridesharing and yellow taxi have the same level of weekly periodicity, ridesharing trips are less periodic in the south areas and more periodic in the north areas of Manhattan than yellow taxi. In the bikesharing system, the comparison between membership and causal trips clearly demonstrates more periodic patterns of membership trips. As bikesharing trip volume is less than ridesharing and yellow taxi, the overall weekly periodicity of bikesharing is stronger than ridesharing and yellow taxi.

To examine time-varying mobility patterns, we analyze bimonthly trip data at an hourly resolution and quantify the weekly periodicity of different travel modes in Manhattan. As reported in Table 2, weekly periodicity is generally lower in the early and late months of 2024 compared to the middle of the year. In particular, mobility regularity in November and December declines remarkably, likely due to winter weather and holiday-related disruptions. While subway trips exhibit the strongest periodicity overall, they also experience a pronounced reduction during these months. By contrast, ridesharing and yellow taxi maintain relatively stable weekly periodicity throughout the year. Bikesharing
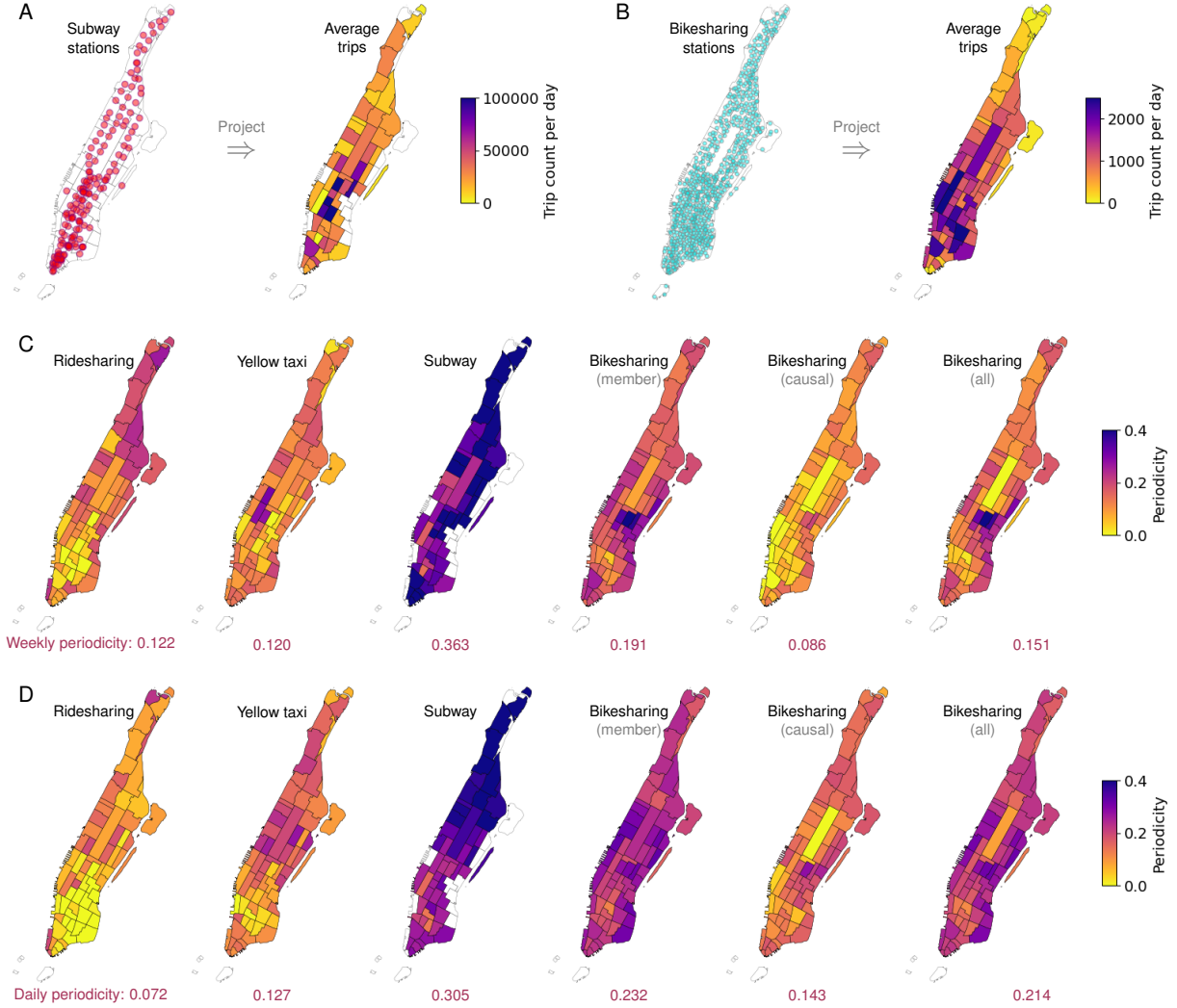
8

Figure 6: Periodicity in multi-modal mobility trip data of 2024 in Manhattan with the multidimensional sparse autoregression method at sparsity level $\tau = 6$. (A) Subway stations are projected onto 52 areas in Manhattan (i.e., 69 areas in total). (B) Bikesharing stations are projected onto 67 areas in Manhattan. (C) Weekly periodicity of four different travel modes, i.e., ridesharing, yellow taxi, subway, and bikesharing. According to the average periodicity, these travel modes are sorted as subway, bikesharing, ridesharing, and taxi in decreasing order. (D) Daily periodicity in weekday's multi-modal mobility trip data.

demonstrates the greatest seasonal sensitivity, with notably weaker periodicity in colder months (January–April and November–December).

Table 2: Weekly periodicity of bimonthly trip data of four travel modes in Manhattan of 2024.

	Ridesharing	Yellow taxi	Subway	Bikesharing (member)
January & February	0.142	0.144	0.231	0.131
March & April	0.191	0.198	0.449	0.134
May & June	0.152	0.143	0.289	0.216
July & August	0.183	0.133	0.408	0.219
September & October	0.205	0.195	0.415	0.290
November & December	0.046	0.055	0.136	0.120

9

In addition, we consider analyzing the daily periodicity of weekday's trips in Manhattan. The mobility tensor is of size $69 \times 6 \times 6288$ with the temporal dimension spanning 6,288 hours of 262 weekdays in 2024. The support set in our multidimensional sparse autoregression method is optimized as $\Omega = \{1, 23, 24, 95, 96, 120\}$ with sparsity level $\tau = 6$ where the index $k = 24$ refers to the daily cycle. Observing the daily periodicity visualized in Fig. 6D, subway trips are more periodic than other travel modes, but the daily periodicity of the north areas is consistently stronger than the south areas of Manhattan. Compared to the consistent weekly periodicity of ridesharing and yellow taxi trips in Fig. 6C, the daily periodicity of yellow taxi trips is stronger than ridesharing. The daily periodicity of causal bikesharing trips is lower than membership trips, but the distinct regularity became a little marginal.

6   Discussion

Understanding the periodicity of human mobility is critical for revealing the underlying structure of individual and collective movement patterns. Periodic patterns in human mobility—ranging from daily commutes to weekly rituals—are manifestations of deeply rooted behavioral rhythms. These patterns are essential components for making informed decisions, developing effective urban policies, and building more resilient and responsive urban systems. In literature, such rhythms have been extensively studied in chronobiology, behavioral sciences, and temporal network theory [Roenneberg et al., 2004, Dijksterhuis et al., 2006, Holme and Saramäki, 2012], where periodicity reveals regularities in decision-making, attention cycles, and social synchronization. Drawing on these insights, we propose a novel method for quantifying and leveraging periodicity in large-scale mobility data. This enables a deeper understanding of urban rhythms and enhances predictive modeling for transportation planning, epidemic spread [Balcan et al., 2009, Belik et al., 2011, Castillo-Chavez et al., 2016], and urban resilience assessments [Xu et al., 2025].

Our results advance people's understanding of real-world urban systems and show the capability of multidimensional sparse autoregression for quantifying mobility regularity from time series data. Our findings concerning the weekly periodicity—a comparable metric across different time series, spatial areas, and time periods—in NYC and Chicago reveal the disruptive impact of the COVID-19 pandemic on mobility regularity and subsequent recovery trends. In Manhattan, the analysis of daily and weekly periodicity patterns of four different travel modes, including ridesharing, yellow taxi, subway, and bikesharing, demonstrate the rhythm and regularity strength across different spatial areas and bimonthly time phases. The interpretability offered by the multidimensional sparse autoregression is a key advantage, as sparse and non-negative auto-correlations provide actionable insights into the specific temporal lags driving the observed periodicity [Chen et al., 2025b]. The method we present allows substantial analysis of time series data that collected from real-world systems and demonstrated periodicity or seasonality. Despite identifying weekly periodicity, the revealed dominant auto-correlations in the support set enable us to introduce proper differencing operations—a simple yet effective solution to address the non-stationarity issue [Chen et al., 2025c].

As the field of human mobility research moves from analytical studies to data-driven experiments designed to directly learn underlying patterns, there is a growing need to formulate the periodicity quantification problem as a sparse identification of dominant positive coefficients in multidimensional time series autoregression. We have developed an interpretable machine learning method for identifying spatial patterns and temporal dynamics that relate to weekly periodicity in urban areas. While this study focuses on metro passenger flow and ridesharing trips in specific cities, a potential limitation is the focus on these modes. Future work could integrate data from other transportation modes, such as public transit, to capture more comprehensive multi-modal behaviors. In terms of potential applications, this interpretable framework provides a robust foundation for discovering and quantifying temporal patterns related to periodicity and seasonality in complicated data and systems beyond human mobility, including climate systems and healthcare time series.

References

Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779–782, 2008.

Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.

Yosef Sheffi. *Urban transportation networks*, volume 6. Prentice-Hall, Englewood Cliffs, NJ, 1985.

Valérie Guihaire and Jin-Kao Hao. Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, 42(10):1251–1273, 2008.

Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the national academy of sciences*, 106(51):21484–21489, 2009.

Vitaly Belik, Theo Geisel, and Dirk Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X*, 1(1):011001, 2011.

Carlos Castillo-Chavez, Derdei Bichara, and Benjamin R Morin. Perspectives on the role of mobility, behavior, and time scales in the spread of diseases. *Proceedings of the National Academy of Sciences*, 113(51):14582–14588, 2016.

Weiyu Li, Qi Wang, Yuanyuan Liu, Mario L Small, and Jianxi Gao. A spatiotemporal decay model of human mobility when facing large-scale crises. *Proceedings of the National Academy of Sciences*, 119(33):e2203042119, 2022.

Zhanwei Du, Spencer J Fox, Petter Holme, Jiming Liu, Alison P Galvani, and Lauren Ancel Meyers. Periodicity in movement patterns shapes epidemic risk in urban environments. *arXiv preprint arXiv:1809.05203*, 2018.

Naoki Katoh, Akiyoshi Shioura, and Toshihide Ibaraki. Resource allocation problems. *Handbook of combinatorial optimization*, pages 1–93, 2024.

Xinyu Chen and Lijun Sun. Bayesian temporal factorization for multidimensional time series prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4659–4673, 2022.

Peter Midgley. The role of smart bike-sharing systems in urban mobility. *Journeys*, 2(1):23–31, 2009.

Cláudia A Soares Machado, Nicolas Patrick Marie de Salles Hue, Fernando Tobal Berssaneti, and José Alberto Quintanilha. An overview of shared mobility. *Sustainability*, 10(12):4342, 2018.

Alejandro Tirachini. Ride-hailing, travel behaviour and sustainable mobility: an international review. *Transportation*, 47(4):2011–2047, 2020.

Pierre Nouvellet, Sangeeta Bhatia, Anne Cori, Kylie EC Ainslie, Marc Baguelin, Samir Bhatt, Adhiratha Boonyasiri, Nicholas F Brazeau, Lorenzo Cattarino, Laura V Cooper, et al. Reduction in mobility and covid-19 transmission. *Nature communications*, 12(1):1090, 2021.

Michael Batty, Kay W Axhausen, Fosca Giannotti, Alexei Pozdnoukhov, Armando Bazzani, Monica Wachowicz, Georgios Ouzounis, and Yuval Portugali. Smart cities of the future. *The European Physical Journal Special Topics*, 214:481–518, 2012.

Shan Jiang, Joseph Ferreira, and Marta C Gonzalez. Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. *IEEE Transactions on Big Data*, 3(2):208–219, 2017.

Laura Alessandretti, Ulf Aslak, and Sune Lehmann. The scales of human mobility. *Nature*, 587(7834):402–407, 2020.

Zhenhui Li, Bolin Ding, Jiawei Han, Roland Kays, and Peter Nye. Mining periodic behaviors for moving objects. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1099–1108, 2010.

Bhaskar Prabhala, Jingjing Wang, Budhaditya Deb, Thomas La Porta, and Jiawei Han. Leveraging periodicity in human mobility for next place prediction. In *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 2665–2670. IEEE, 2014.

Gabriel Goulet-Langlois, Haris N Koutsopoulos, Zhan Zhao, and Jinhua Zhao. Measuring regularity of individual travel patterns. *IEEE Transactions on Intelligent Transportation Systems*, 19(5):1583–1592, 2017.

Yourong Huang, Zhu Xiao, Dong Wang, Hongbo Jiang, and Di Wu. Exploring individual travel patterns across private car trajectory data. *IEEE Transactions on Intelligent Transportation Systems*, 21(12):5036–5050, 2019.

Douglas do Couto Teixeira, Jussara M Almeida, and Aline Carneiro Viana. On estimating the predictability of human mobility: the role of routine. *EPJ Data Science*, 10(1):49, 2021.

Douglas Do Couto Teixeira, Aline Carneiro Viana, Jussara M Almeida, and Mrio S Alvim. The impact of stationarity, regularity, and context on the predictability of individual human mobility. *ACM Transactions on Spatial Algorithms and Systems*, 7(4):1–24, 2021.

Shuo Tao, Jingang Jiang, Defu Lian, Kai Zheng, and Enhong Chen. Predicting human mobility with reinforcement-learning-based long-term periodicity modeling. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12 (6):1–23, 2021.

Ed Manley, Chen Zhong, and Michael Batty. Spatiotemporal variation in travel regularity through transit user profiling. *Transportation*, 45:703–732, 2018.

Chen Zhong, Ed Manley, Stefan Mueller Arisona, Michael Batty, and Gerhard Schmitt. Measuring variability of mobility patterns from multiday smart-card data. *Journal of Computational Science*, 9:125–130, 2015.

Liye Zhang and Jie Song. The periodicity and initial evolution of micro-mobility systems: A case study of the docked bike-sharing system in new york city, usa. *European transport research review*, 14(1):27, 2022.

Steven L Brunton and J Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2022.

Eduardo Mucelli Rezende Oliveira, Aline Carneiro Viana, Carlos Sarraute, Jorge Brea, and Ignacio Alvarez-Hamelin. On the regularity of human mobility. *Pervasive and Mobile Computing*, 33:73–90, 2016.

Xinyu Chen, Vassilis Digalakis Jr, Lijun Ding, Dingyi Zhuang, and Jinhua Zhao. Interpretable time series autoregression for periodicity quantification. *arXiv preprint arXiv:2506.22895*, 2025a.

Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.

Federico Delussu, Michele Tizzoni, and Laetitia Gauvin. The limits of human mobility traces to predict the spread of covid-19: A transfer entropy approach. *PNAS nexus*, 2(10):pgad302, 2023.

W James Murdoch, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu. Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080, 2019.

Cynthia Rudin, Chaofan Chen, Zhi Chen, Haiyang Huang, Lesia Semenova, and Chudi Zhong. Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistic Surveys*, 16:1–85, 2022.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

Rodolphe Jenatton, Jean-Yves Audibert, and Francis Bach. Structured variable selection with sparsity-inducing norms. *The Journal of Machine Learning Research*, 12:2777–2824, 2011.

Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

Xinyu Chen, Chengyuan Zhang, Xiaoxu Chen, Nicolas Saunier, and Lijun Sun. Discovering dynamic patterns from spatiotemporal data with time-varying low-rank autoregression. *IEEE Transactions on Knowledge and Data Engineering*, 36(2):504–517, 2024.

James D Hamilton. *Time series analysis*. Princeton university press, 2020.

Dimitris Bertsimas, Angela King, and Rahul Mazumder. Best subset selection via a modern optimization lens. 2016.

Dimitris Bertsimas, Jean Pauphilet, and Bart Van Parys. Sparse regression. *Statistical Science*, 35(4):555–578, 2020.

Dimitris Bertsimas, Vassilis Digalakis Jr, Michael Lingzhi Li, and Omar Skali Lami. Slowly varying regression under sparsity. *Operations Research*, 2024.

Andreas M Tillmann, Daniel Bienstock, Andrea Lodi, and Alexandra Schwartz. Cardinality minimization, constraints, and regularization: a survey. *SIAM Review*, 66(3):403–477, 2024.

Till Roenneberg, Tim Kuehnle, Peter P Pramstaller, Jan Ricken, Miriam Havel, Angelika Guth, and Martha Merrow. A marker for the end of adolescence. *Current biology*, 14(24):R1038–R1039, 2004.

Ap Dijksterhuis, Maarten W Bos, Loran F Nordgren, and Rick B Van Baaren. On making the right choice: The deliberation-without-attention effect. *Science*, 311(5763):1005–1007, 2006.

Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.

Fengli Xu, Qi Wang, Esteban Moro, Lin Chen, Arianna Salazar Miranda, Marta C González, Michele Tizzoni, Chaoming Song, Carlo Ratti, Luis Bettencourt, et al. Using human mobility data to quantify experienced urban inequalities. *Nature Human Behaviour*, pages 1–11, 2025.

Xinyu Chen, HanQin Cai, Fuqiang Liu, and Jinhua Zhao. Correlating time series with interpretable convolutional kernels. *IEEE Transactions on Knowledge and Data Engineering*, 37(6):3272–3283, 2025b.

Xinyu Chen, Chengyuan Zhang, Xi-Le Zhao, Nicolas Saunier, and Lijun Sun. Forecasting sparse movement speed of urban road networks with nonstationary temporal matrix factorization. *Transportation Science*, 2025c.

A Data, Material, and Code Availability

Data and code of this study will be available at `https://github.com/xinychen/integers`. For an intuitive understanding of time series periodicity, we also provide an interactive visualization tool on the ridesharing trip time series.

B Methodology

While the optimization problem we have posed has the potential to learn a significant amount of information about periodicity in time series data, estimating the coefficient vectors $\{w_{n,\gamma,k}\}_{n\in[N],\gamma\in[\delta],k\in[d]}$ of N spatial areas and δ phases when N or δ become even moderately large becomes extremely computationally expensive. Thus, instead of estimating individual coefficient vectors $\{w_{n,\gamma,k}\}_{n\in[N],\gamma\in[\delta],k\in[d]}$, we first estimate global auto-correlations with sparsity constraints. The optimization problem for estimating a global coefficient vector $\boldsymbol{w} = \{w_k\}_{k\in[d]} \in \mathbb{R}^d$ from the human mobility data $\{x_{n,\gamma,t}\}_{n\in[N],\gamma\in[\delta],t\in[T_\gamma]}$ is given by

$$\min_{\{w_k\},\Omega} \sum_{n=1}^{N}\sum_{\gamma=1}^{\delta}\sum_{t=d+1}^{T_\gamma} \left(x_{n,\gamma,t} - \sum_{k\in\Omega} w_k x_{n,\gamma,t-k} \right)^2$$

$$\text{subject to} \begin{cases} |\Omega| \leq \tau,\ \tau \in \mathbb{Z}^+, \\ w_k \geq 0,\ \forall k \in \Omega, \\ w_k = 0,\ \forall k \notin \Omega, \\ \sum_{k\in\Omega} w_k = 1, \end{cases}$$

which can be solved by mixed-integer optimization. The support set is therefore optimized Ω simultaneously.

Given the global support set Ω, learning the individual coefficients $\{w_{n,\gamma,k}\}_{n\in[N],\gamma\in[\delta],k\in[d]}$ can be simplified as follows,

$$\min_{\{w_{n,\gamma,k}\}} \sum_{t=d+1}^{T_\gamma} \left(x_{n,\gamma,t} - \sum_{k\in\Omega} w_{n,\gamma,k} x_{n,\gamma,t-k} \right)^2$$

$$\text{subject to} \begin{cases} w_k \geq 0,\ \forall k \in \Omega, \\ w_k = 0,\ \forall k \notin \Omega, \\ \sum_{k\in\Omega} w_k = 1, \end{cases}$$

for any $n \in [N]$ and $\gamma \in [\delta]$. We solve this optimization problem by using quadratic optimization.

C Metro Passenger Flow Periodicity

C.1 Weekly periodicity on the 30-minute passenger flow time series data

Fig. 7 shows the weekly periodicity across 80 stations by using the proposed model. While this study discusses passenger flow time series with both hourly and 30-minute time resolutions, it is possible to see the periodicity pattern shift of passenger flow from hourly data to 30-minute data. On these time series, the sparsity level of the proposed model is set as $\tau = 4$, and the support set is optimized as $\Omega = \{1, 48, 335, 336\}$ (i.e., covering both daily cycle $k = 48$ and weekly cycle $k = 336$). Thus, the weekly periodicity is revealed by the auto-correlation at the index $k = 336$. Comparing to the hourly time series, the weekly periodicity in Fig. 7 is relatively small because 30-minute time series demonstrate stronger local auto-correlations than hourly time series. Of the results, the inflow is more periodic than the outflow in the end stations, while the inflow is less periodic than the outflow in downtown areas. Thus, the time resolution impacts the time series periodicity quantification. Since the periodicity is a relative value, the comparison among time series with the same time resolution is still meaningful.

D Regularity of Ridesharing Trips

D.1 Weekly periodicity of ridesharing pickup trips with different sparsity levels

In this study, we consider the sparsity level on the NYC ridesharing data as $\tau = 6$, covering weekly cycle in the support set. For comparisons, we set the sparsity level of the proposed model as $\tau = 4$ in which the support set is optimized
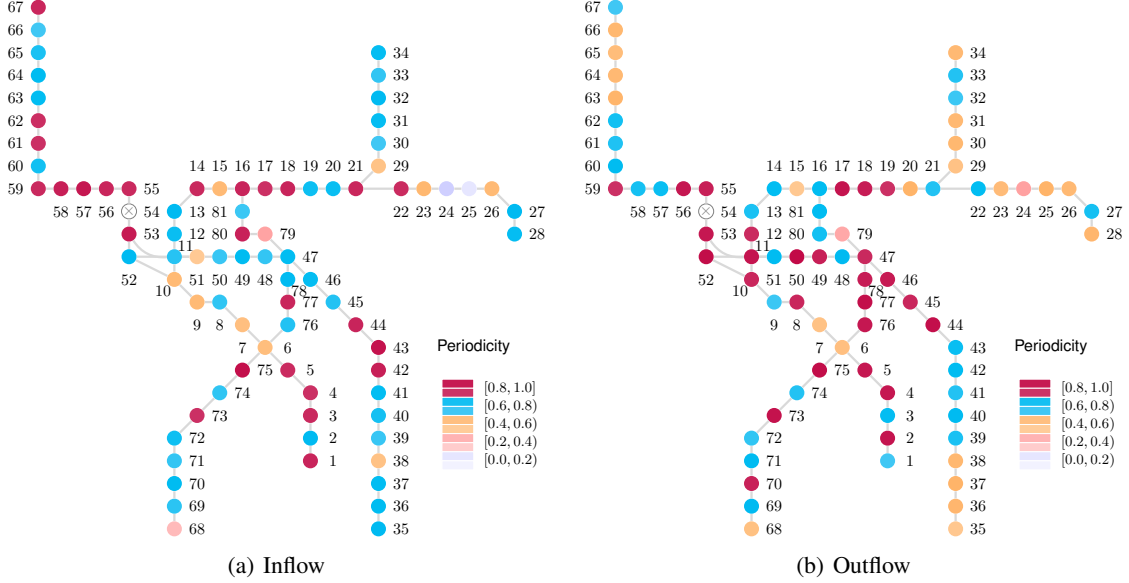
(a) Inflow   (b) Outflow

Figure 7: Weekly periodicity of 30-minute inflow and outflow passenger flow in the Hangzhou metro system.

as $\Omega = \{1, 23, 167, 168\}$. As shown in Fig. 8, although the weekly periodicity at the sparsity level $\tau = 4$ is a little different from $\tau = 6$, we can also observe the reduction of weekly periodicity in 2020 due to the COVID-19 pandemic, following by the recovery of human mobility regularity to the normal status. Thus, the interpretability of weekly periodicity is robust to the setting of different sparsity levels.
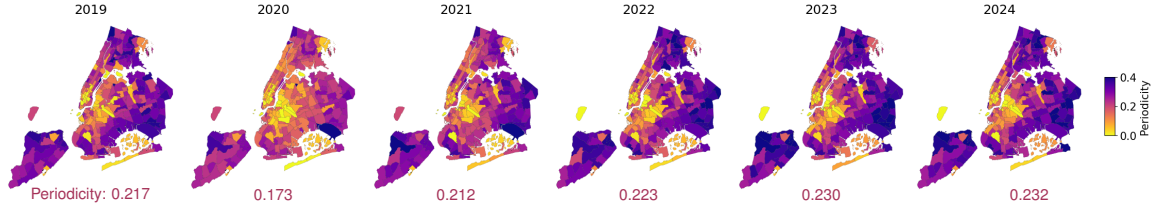

Figure 8: Weekly periodicity of ridesharing pickup trips quantified by the proposed model with sparsity level $\tau = 4$.

D.2   Weekly periodicity of ridesharing dropoff trips with different sparsity levels

In terms of dropoff trips, we set the sparsity levels of the proposed model as $\tau = 4, 6$ as follows.

- The support set with sparsity level $\tau = 4$ is $\Omega = \{1, 23, 167, 168\}$.
- The support set with sparsity level $\tau = 6$ is $\Omega = \{1, 23, 24, 143, 167, 168\}$.

As shown in Fig. 9, the dropoff trips of NYC ridesharing data demonstrated similar temporal patterns of weekly periodicity as the pickup trips. Trips with destination to downtown areas are less periodic than suburban areas. In 2020, a remarkable reduction of weekly periodicity can be clearly observed. In 2022, the weekly periodicity recovered to the same level as 2019. These findings intuitively verify the impact of the COVID-19 pandemic on mobility regularity in urban areas.
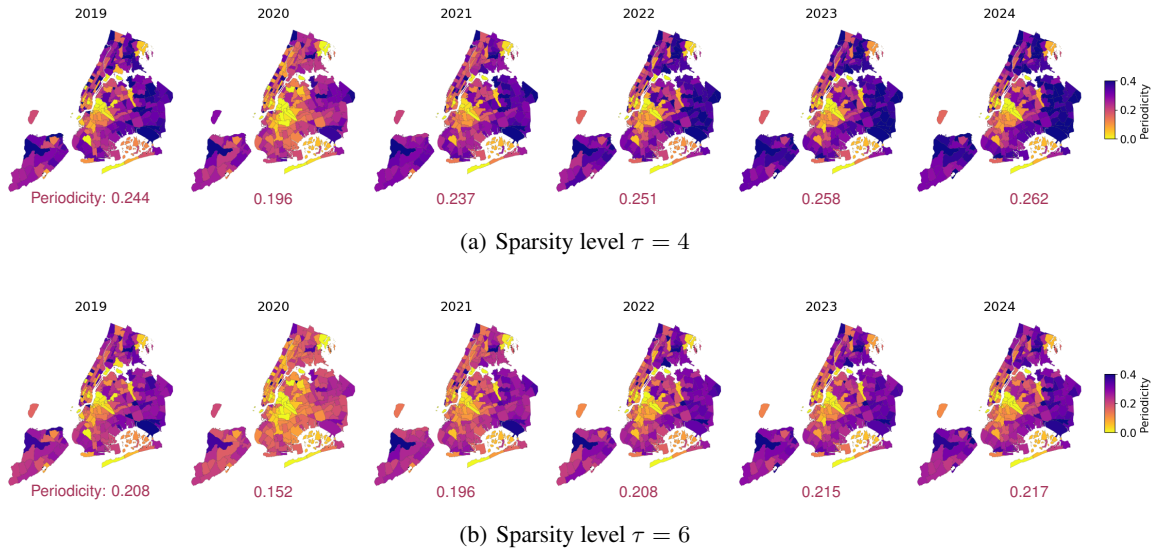
14

(a) Sparsity level $\tau = 4$

(b) Sparsity level $\tau = 6$

Figure 9: Weekly periodicity of ridesharing dropoff trips quantified by the proposed model with sparsity levels $\tau = 4, 6$.