EXTRACTING SPATIO-TEMPORAL PATTERNS OF

TAXI TRIPS DURING COVID-19:

A CASE STUDY OF CHICAGO

by

Ben Wang, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Geography
August 2021

Committee Members:

Yihong Yuan, Chair

F. Benjamin Zhan

Rasim M. Musal

ACKNOWLEDGEMENTS

I would like to first thank my thesis committee. Dr. Yihong Yuan, as my advisor, gave me invaluable supports with her expertise during the research design and thesis writing process. Dr. F. Benjamin Zhan and Dr. Rasim Musal also provided helpful feedback and have been great teachers and mentors. They prepared me well to get to this place in my academic life. This project would not be nearly as successful and rewarding without their help.

I would especially like to thank my family. My girlfriend, Yaping, has been extremely supportive of me throughout this entire process to help me get to this point. My parents deserve special thanks for their continued support and encouragement. Without such a team behind me, I doubt that I would be in this place today.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

Abbreviation	**Description**
ARIMA	Autoregressive integrated moving average
DTW	Dynamic time wrapping
FCD	Floating car data
ICTs	Information communication technologies
LBSM	Location-Based Social Media
LOESS	Locally estimated scatterplot smoothing
LSTM	Long short-term memory
O-D	Origin-destination
SEATS	Seasonal Extraction in ARIMA Time Series
STL	Seasonal-Trend decomposition using LOESS

ABSTRACT

In 2020, the whole world experienced unprecedented challenges caused by COVID-19, and the United States is one of the most impacted areas. With the high contagiousness of COVID-19, human mobility became the catalyst of the pandemic. The rapid growth of big geo-data also provides an opportunity and a challenge to explore human mobility during a pandemic. This research extracts spatio-temporal urban dynamics from floating car data (FCD) in Chicago during and before COVID-19. The taxi trip records are aggregated by the hour at the community area level, so there is a taxi trip time series in each community area. We applied a time series decomposition method, Seasonal-Trend decomposition using LOESS (STL), to analyze taxi trips' spatio-temporal patterns. STL can divide the original time series into different components, including trend, seasonality, and residuals. We also clustered the trend and seasonal effects of time series in different community areas of Chicago. The results show that time series decomposition is useful for extracting short-term and long-term patterns in urban dynamics. The comparison of the spatio-temporal patterns of taxi trips before and during COVID-19 indicates a sharp decrease and less diversity of human mobility during a pandemic.

1 INTRODUCTION

In 2020, the COVID-19 pandemic became a global crisis in many countries, causing devastating repercussions on public health and social stability (Oldekop et al. 2020). The United States was also experiencing unprecedented challenges with a large number of confirmed cases and deaths (World Health Organization 2020). The State of Illinois was one of the states where earlier cases were found in January, but it did not order to close public places until March 13th. A stay-at-home request was initiated on March 20th - almost two months after the first confirmed case in Illinois. In addition, on May 5th, the State of Illinois started a plan to reopen the state for economic resumption (Chicago Tribune 2020). By June 1st, there were 47,278 confirmed cases in the City of Chicago (City of Chicago 2020).

With the high contagiousness of COVID-19, human mobility became the catalyst of the pandemic. This has been a topic of interest since humans have been traveling from one place to another for all of the time (Rodrigue 2016). Nowadays, people's usage of transportation remarkably increases their mobility, and data about their travel activities have been generated (Song, Kanasugi, and Shibasaki 2016). In epidemiology, research focusing on human mobility aims to provide a better understanding of the roles of human mobility and decision making in influencing the spreading of the epidemic through transportation restrictions, individual quarantine, and source tracking (Bajardi et al. 2011; Oliver et al. 2020). With the rapid development of geographic information systems, computer science, and information communication technologies (ICTs), geographic information produced by ubiquitous electronic devices such as smartphones and sensors in vehicles has become a new source of big data (Moreira-Matias et al. 2013). It is both an opportunity and a challenge to explore how big data with geographic information (aka big geo-data) can help analyze human mobility during an epidemic or a pandemic.

The collection of big geo-data relies on the prevalence of ICTs, which implement communication devices to provide big data sources, including mobile phone data, Bluetooth data, Location-Based Social Media (LBSM) data, etc. To monitor the spread of COVID-19 during the Spring Festival travel rush in China, high-tech companies like Baidu collected real-time mobile phone data to help analyze the transmission of the coronavirus in interstate traveling (Kraemer et al. 2020). In addition, the Chinese government launched a policy called the Health Code system to monitor individual health conditions in their mobile phones, which is also vital to reduce the transmission of COVID-19 (Pan 2020). ICTs also use built-in communication devices in vehicles and collecting big geo-data, such as floating car data (FCD), to improve the transportation system (Croce et al. 2019). Unlike traditional traffic data collected by fixed devices, FCD contains timestamps and locations collected by moving vehicles (Wikipedia 2020). In urban planning and transportation analysis, FCD helps to extract land use and urban structure (Liu et al. 2012b; Xi Liu et al. 2015), reduce traffic congestion, and predict traffic conditions to improve the efficiency of transportation (Xu, Yue, and Li 2013; De Fabritiis, Ragona, and Valenti 2008). While there is much research with potential applications on pandemics using mobile phone data, less research focuses on FCD for the following reasons: first, the main mode of COVID-19 spreading is person-to-person, but the interaction and protection measures between passengers and the driver are unknown through FCD. In a public vehicle like a taxi, the virus is more likely to be spread by frequently touched surfaces (CDC 2020a; CDC 2020b). Mobile phone data can help identify highly populated areas, which is a fundamental application in understanding virus transmission (Fang et al. 2020). Therefore, the FCD analysis is an indirect approach to reflect the transmission process, making it not the first choice when the mobile phone data is available. Second, COVID-19 is a global pandemic, and the transmission process is multi-scale in geography, while the FCD is usually collected on the local scale. Third, for previous studies in epidemiology, such as SARS and H1N1, few datasets are accessible.

Although FCD has certain limitations, it can be a useful data source to study COVID-19 at an urban scale in the United States. For example, FCD can provide a direct measure for traffic conditions and shows more advantages than mobile phone data in data availability. Due to data privacy, most public mobile phone data for tracking mobility in COVID-19 are aggregated (Gao et al. 2020), which is difficult to extract intra-urban patterns at a finer scale. In contrast, FCD, like the taxi dataset from the City of Chicago, contains each record with origin-destination (O-D) information, which is at a finer scale. These datasets often contain millions of trip records and are accessible in many cities in the United States, such as Chicago and New York. Additionally, since airline travels decreased significantly during COVID-19 (Wallace 2020), a large proportion of COVID-19 transmission was caused by human mobility at an intra-urban scale, which can be feasibly measured by FCD. Comparing other types of FCD, taxi data is more available than private car data and rideshare service data, and it can be tied to the mobility patterns of high-risk groups as one taxi will serve many passengers in a day. As a result, there is a greater chance of exposure to COVID-19 than driving private cars. Therefore, in this study, we are interested in how the taxi dataset in the City of Chicago can reveal human mobility patterns during COVID-19.

Analyzing the time series in FCD can help to extract spatio-temporal patterns, which are essential in understanding human mobility. One method in time series analysis is decomposition, which decomposes the time series into different components: trend, seasonality, and residuals. In physical geography, time series decomposition is a frequently used method to analyze abnormal weather and shows excellent performance (Deng 2014). In transportation research, this decomposition method is also used for data preprocessing and special event detection (Zhu and Guo 2017). In this study, we explored how to use time series decomposition as the primary method to extract the temporal dynamic of human mobility on a finer scale. In addition, this study compared the spatio-temporal patterns of human mobility before and during a pandemic.

In summary, this study investigated the spatio-temporal patterns of taxi data in Chicago during COVID-19 using time series decomposition as well as a clustering analysis. With the results of the time series analysis and clustering analysis, this study visualized and examined the spatio-temporal patterns and movements in human mobility. This study using FCD extracted patterns in human mobility from the regularity of the time series and demonstrated the uncertainty in human mobility by the high ingroup variance in the time series clusters. With a massive dataset of FCD available, this study investigated how to use time series decomposition to extract meaningful temporal patterns before and during COVID-19 and explore the differences between urban functional regions based on these temporal patterns. The outcomes of this study help to improve the efficiency of public transportation and build a better understanding of human mobility, contributing to urban analysis and planning. Furthermore, by comparing human mobility before and during COVID-19, it can help decision-making related to transportation and human mobility in the pandemic.

2 PROBLEM STATEMENT

The objective of this research is to investigate the spatio-temporal patterns of FCD using time series decomposition as well as a clustering analysis to analyze human mobility during COVID-19. With high-frequency taxi trip records in FCD, this study can examine the performance of time series decomposition in extracting patterns from time series. This study uses a time series decomposition method based on Locally estimated scatterplot smoothing (LOESS), named Seasonal-Trend decomposition using LOESS (STL), on the time series in FCD. STL has advantages in extracting seasonality of a short period finer than a quarter or a month. As FCD in this study has a high sampling rate, STL can explore the details in the time series of FCD.

Furthermore, due to the spread of COVID-19 and the enforced social distancing orders, the patterns of taxi trips in Spring 2020 in Chicago can be very different from the patterns from the same months in 2019. Understanding abnormal patterns of taxi trips will help analyze characteristics of human mobility in the pandemic and improve the decision-making process related to human mobility and public transportation to control the spread of the virus better.

Therefore, this research aims to answer these two questions:

1. Methodologically, how can time series decomposition analyze the temporal patterns of taxi data in Chicago?
2. Empirically, how are the patterns different before and during COVID-19?

The following sections include a literature review on previous human mobility studies, the methodology of this research, and results and discussions.

3 LITERATURE REVIEW

3.1 Human mobility studies in the big data era

With the increasing accessibility to big geo-data, human mobility has increased awareness in the fields of transportation, urban planning, and data science (Eagle and Pentland 2006). Although studies with small data like surveys and self-reports improved the understanding of human mobility, it has weaknesses when investigating multi-level data that combines data varying from a local scale to a global scale. Previous studies found that individual human mobility follows a certain spatio-temporal regularity and has a spatial probability distribution (Gonzalez, Hidalgo, and Barabasi 2008), but big data research indicates that the nature of human mobility is more complicated than certain patterns. Analyzing big geo-data provides more opportunities to investigate human mobility patterns, factors that impact travel behaviors (e.g., travel purposes, environment effects, and socio-demographics), and develop models to predict human travel behaviors (Chen et al. 2016). Previous studies introduced the concept of social sensing to integrate big geo-data into the analysis of the social environment, which records human mobility patterns, spatial interaction, and place semantics to extract socioeconomic patterns (Yu Liu et al. 2015). Recently, with the development of machine learning and deep learning, more advanced analytical methods are also introduced into big geo-data studies. Long short-term memory (LSTM), Levy flight, and random walk models are the most popular approaches to analyze transportation patterns and predict mobility behaviors (Song, Kanasugi, and Shibasaki 2016; Zhu and Laptev 2017; Xu et al. 2017). These studies mostly developed machine learning models for traffic demand forecasting, which is one of the most crucial topics in transportation studies.

Human mobility not only reflects the movement patterns of individuals but also provides an alternative data source to investigate the characteristics of urban spaces (Andris 2016). Previous studies analyzed human mobility at both the individual level and urban level and found that the human mobility patterns changed gradually along with the development of urban structures (Gonzalez, Hidalgo, and Barabasi 2008; Kang et al. 2012). At an individual level, understanding human mobility helps to extract information about individual activity spaces and predict individual movements. At an urban level, it helps to identify urban land uses and dynamic social activities more effectively than survey data with limited sample size or remote sensing data that cannot fully capture the dynamics of human mobility. Human mobility can also influence the pandemic by helping or controlling the spread of the virus at the urban level (Belik, Geisel, and Brockmann 2011). The result of epidemic prevention work is determined by the contribution of the state-wide or city-wide orders because there was no uniform national action in the United States, but many cities or states responded to control the pandemic with their own solutions during COVID-19 (Haffajee and Mello 2020). An analysis of human mobility at the urban level could improve the process of prevention work by optimizing the distribution of COVID-19 testing, mobile cabin hospitals, and other medical resources and services (Weissman et al. 2020).

For collecting data on dynamic human mobility, there are several typical datasets to obtain sufficient big geo-data: mobile phone dataset, which tracks human movements directly; Bluetooth dataset, which can monitor the number of people nearby; location-based social media (LBSM), which contains abundant text information with geotags; and FCD, which can monitor human mobility along with a street network. Comparing other datasets, collecting FCD is an efficient and low-cost approach to gather information about traffic intensity and distribution, which contains human mobility patterns in the street network (Jiang, Yin, and Zhao 2009). The taxi dataset is a kind of FCD, which is often more accessible than personal car data or ride share service data from private companies. Meaningful patterns from FCD can be extracted by data mining methods like classification and clustering for further analysis (Mazimpaka and Timpf 2016). Another important application of FCD is to provide a better understanding of urban structures by investigating the interdependence between land uses and traffic (Liu, et al. 2012b). In addition, as the COVID-19 pandemic had aggregated impacts on urban transportation (Gupta et al. 2020), FCD studies can help provide a better understanding of human mobility in the epidemic at the urban level.

3.2 Using FCD for Human Mobility Studies

Yuan and Raubal (2016 summarized several kinds of analysis methods on big geo-data in human mobility studies. These analysis methods have different processes based on the spatio-temporal attributes in big geo-data. FCD often include useful attributes about trip O-D pairs, such as the locations and timestamps of pickup/drop-off points and the travel time. These attributes are valuable for investigating travel patterns and urban structures.

Figure 1 shows a popular method to process FCD. First, it separates the urban area into small cells, then extracts the temporal mobility patterns separately. Next, it summarizes, classifies, or clusters the patterns. Finally, it predicts the traffic demands in different areas or connects patterns with urban land uses, like train stations, leisure facilities, and business districts (Qi et al. 2011; Liu et al. 2012a). Based on this idea, previous studies were able to recognize various land-use types like commercial, industrial, residential, institutional, and recreational types and provided a more efficient approach to developing land use data other than traditional surveys or remote sensing images (Liu, et al. 2012b). Another study applied network science methods like the Infomap algorithm as a community detector to reveal sub-regional structures (Liu, et al. 2015). These methods can reflect the polycentric structure and the heterogeneity of urban spaces.

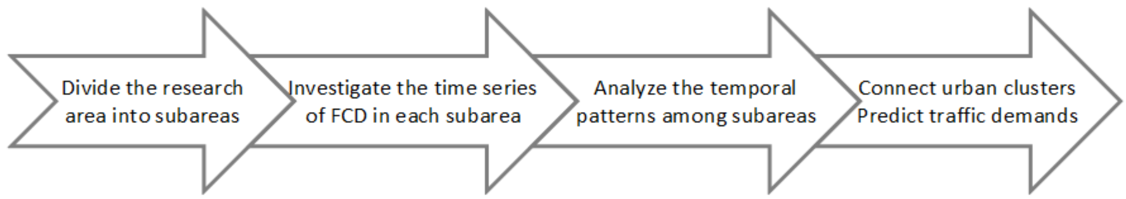


Figure 1. A method to analyze FCD.

The urban dynamics and the integration of multi-land uses make the urban system more complex in transportation, human activities, social functions, and the urban structure (Qi, et al. 2011). Previous studies used the traffic flows as an important indicator for inferring social functions or land uses in various regions of the urban system. However, analyzing only the spatial factor from the traffic flow data is unsubstantial to investigate the multi-faceted social functions and complexity of urban spaces (Goddard 1970). Besides the spatial factors in GIS studies, the time has always been an essential dimension in analyzing geographic phenomena due to the dynamic nature of human mobility and social activities. This dynamic nature is reinforcing the spatial complexity, in which human activities make the social functions spatially and temporally unique (Liu, et al. 2012a; Song, Kanasugi, and Shibasaki 2016).

Investigating the spatio-temporal patterns helps to improve transportation and urban planning (Yuan and Raubal 2016). Related research extracted regular travel time patterns and took advantage of the spatial characteristics of a taxi network, improving the work efficiency of the taxi system (Gao et al. 2013; Moreira-Matias, et al. 2013). Other topics of analyzing spatio-temporal patterns include predicting traffic demands and detecting the abnormality in human mobility or urban events based on time series analysis. These methods of time series analysis vary according to the research purposes. Xu, et al. (2017 used neural networks to build a sequence learning model to predict taxi demands in different areas of New York City. Rodrigues, Markou, and Pereira (2019 combined LBSM data with the New York dataset and used deep learning to examine the performance of multi-source big geo-data analysis on predicting taxi movements. Other studies used dynamic time wrapping (DTW) or time series decomposition to extract abnormal patterns in human mobility (Yuan and Raubal 2012; Zhu and Guo 2017). Especially, Zhu and Guo (2017 used STL to extract the residual component from the taxi time series to analyze the special days or abnormal events in New York City. This research uses the same time series decomposition method, STL, but aims to explore new patterns of human mobility before and during COVID-19.

3.3 Time series decomposition and its applications

Time series analysis is a statistical process to describe the distribution of the observations over time. It is broadly implemented in many fields, such as ecology, economics, and physics (Hyndman and Athanasopoulos 2018). There are many widely used methods for time series analysis, such as classical time series regression, autoregressive integrated moving average (ARIMA) model, spectral analysis, and state-space models. The classical time series regression is based on conventional linear regression and assumes multiple variables are influencing the observation, which is suitable for a fixed period. The ARIMA model combines autoregression and moving average to solve the delayed and linear relations in time series (Hyndman and

Athanasopoulos 2018). The spectral analysis focuses on the frequencies in time series, transforming the sequence to vectors to measure the similarities or variations in time (Koopmans 1995). State-space models recognize the state of the observation at each time point to help forecast states in the future. One type of state-space models is structural models, which investigate the structure and sub-components of the time series, such as trend, seasonality, and residuals. One primary method of structure models is time series decomposition. It is often helpful to split a time series into several components, including trend, seasonality, and the residuals, each representing an underlying component of the characteristics of the times series (Deng 2014). These components reflect different natures of the variations in the time series and are useful in different fields with particular targets, varying from the stock trend to weather forecast and traffic condition. For classic time series decomposition, there are three typical components: trend, seasonality, and residuals. In the time series, the trend reflects a general direction of the observations, the seasonality indicates a regularity in a short period, repeating along the whole period. Residuals contain random fluctuations in the time series, which explains anything out of trend and seasonality. The classic time series decomposition usually uses an additive method to decompose the structure of time series (Hyndman and Athanasopoulos 2018):

$$y_t = S_t + T_t + R_t \quad (1).$$

In equation (1), y_t represent the count of pickups/drop-offs in the time series at time *t* in each community area. In this study, S_t is the seasonal component, T_t is the trend component, and R_t is the remainder component. The additive decomposition is the most appropriate if the scale of the seasonality or the trend will not change proportionally along with the time series.

Besides classical decomposition, other methods include X11 decomposition, Seasonal Extraction in ARIMA Time Series (SEATS), STL, etc. (Shumway and Stoffer 2017). Classical decomposition extracts the trend using moving averages; although it has a simple mathematic process and is easy to compute, it shows weaknesses in analyzing seasonality in a long period, outliers, and will lose data of first few and last few points in the series (Hyndman and Athanasopoulos 2018). Another popular method for time series decomposition is LOESS (STL). The basic structure of STL is a recurrent process to smooth the original time series to a new trend series and extract the seasonality by LOESS. Then it uses the new trend as the input into this repeated process until extracting most variation into trend, seasonality, and residuals (Cleveland 1990). As time series decomposition is a structural model in space-state models, it can efficiently describe the states of the observations, such as the dynamics of human mobility, at each time point by separating the original time series into trend, seasonality, and residuals. Instead of using the original series, these decomposed components can serve as vectors to measure the similarities and classify the states of observations into different groups of fluctuation characteristics by trend and seasonality.

As time series decomposition is applied as a data preprocessing method to identify the structure of times series data, researchers often need further analysis to discover how external factors, such as marketing initiatives and changes in socioeconomic conditions, may influence the fluctuation of time series (Zhu and Guo 2017; Rodrigues, Markou, and Pereira 2019). For big geo-data like FCD, locations and related events are also important factors that can help reveal urban patterns, but time series decomposition does not consider these critical variables. Therefore, the step after decomposition is to extract spatial patterns in FCD by analyzing the spatial heterogeneity in different sub-areas. This study uses clustering analysis as spatial analysis after time series decomposition for extracting spatio-temporal patterns from FCD.

4 METHODOLOGY

This research uses the Chicago taxi dataset, which contains information for each taxi trip in Chicago published by the City of Chicago (Chicago Data Portal 2016). According to the FCD processing method discussed in the literature review, the whole city area is divided into sub-areas (i.e., community areas, which are regional areas determined by the City of Chicago based on the neighborhoods for census and urban planning purposes). Each community area includes aggregated numbers of pickups/drop-offs in a fixed time interval of a day, so we can apply time series decomposition to extract temporal patterns. After that, we apply a clustering analysis to measure the similarities of these temporal patterns and cluster community areas into different groups. The anticipated temporal patterns include daily and weekly fluctuations of the numbers of taxi trips. These fluctuations reflect short-term and long-term urban dynamics as they coincide with life and work schedules. This research combines the daily and weekly fluctuations to investigate their impact on the clustering results. We also compare the differences in human mobility before and during COVID-19, including time series, clustering results, and urban land use connections.

4.1 Data Source and Preparation

The taxi trip records are collected from the Chicago Data Portal (Chicago Data Portal 2016). Each record includes attributes such as the id of trips and drivers, the timestamps and locations of pickup and drop-off events, etc. The timestamps are rounded to the nearest 15 minutes, and locations are masked at the census tract level or the community area level if there are fewer than three records in the census tract in a 15-minutes time period (Chicago Data Portal 2016).

This research used two periods of data: March 2019 (1.5 million records) and March 2020 (0.5 million records) to compare the difference in the taxi trips on a yearly basis before and during COVID-19. Figure 2 shows the spatial distribution of taxi pickup locations. Many records in the central area are aggregated at the census tract level, while in the outskirts, most records are at the community area level.

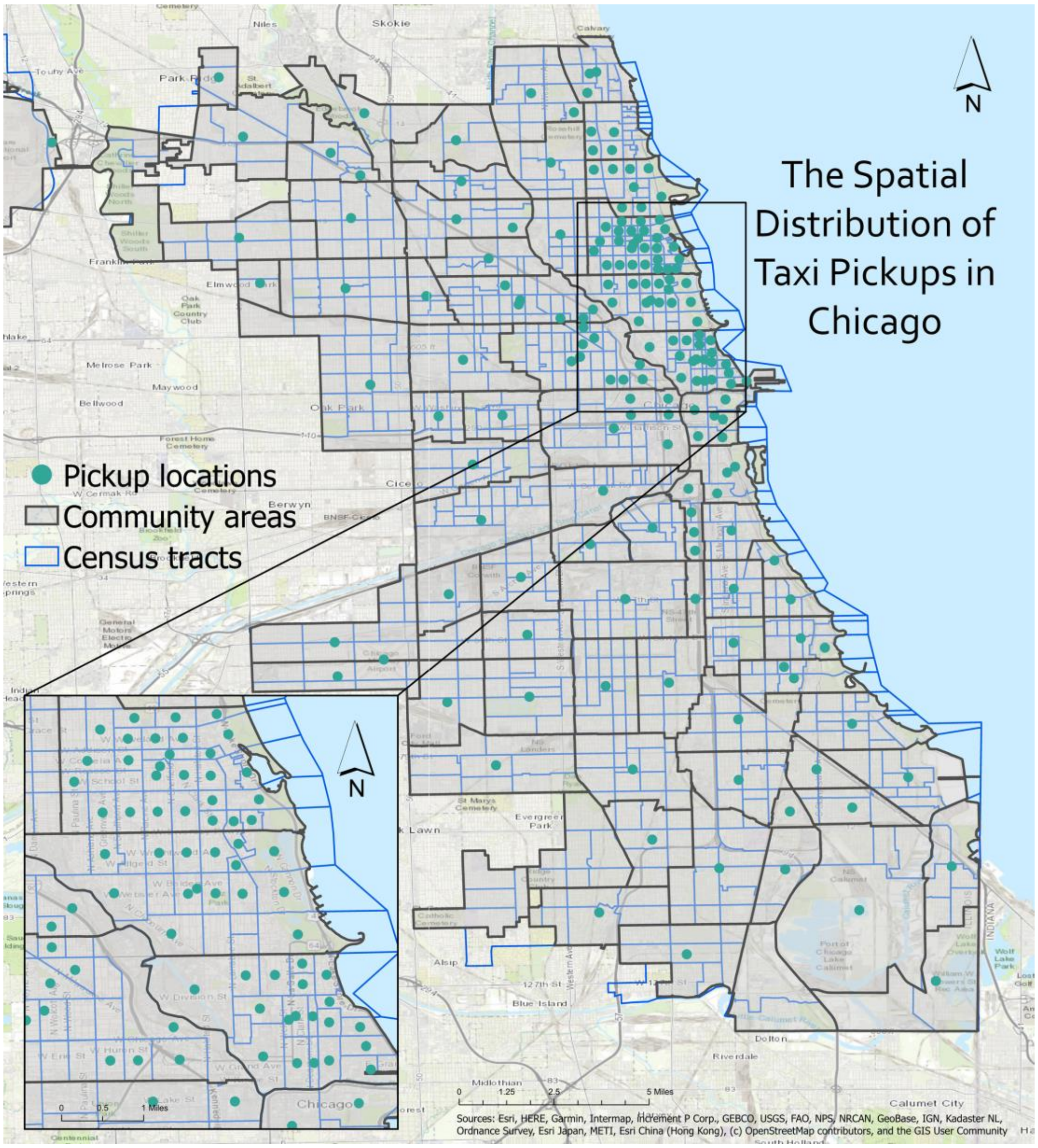


Figure 2. The map of the taxi pickup distribution in Chicago.

This research first divides the study area into community areas and aggregates records based on a fixed temporal granularity. We set the granularity as 1 hour because it can retain the detailed hourly patterns in a day while reducing noise in the original time series with a 15-minute resolution. Figure 3 visualizes the time series of taxi pickups in two community areas after aggregation. The original time series contains fluctuations of the trend, seasonality, and residual components. As can be seen, the range of taxi pickups is different among community areas. The different ranges of taxi pickup numbers will make it difficult to compare fluctuation patterns among community areas in the next clustering analysis. To reduce the range effect, this research uses a normalization process to map the records to the range [0, 1] with the equation:

$$x^* = \frac{x - min}{max - min} \quad (2).$$

In equation (2), x is the number of records during an hour, x^* is the normalized value, and *min/ max* are the numbers of records during the hour with the least/most records in the whole month.

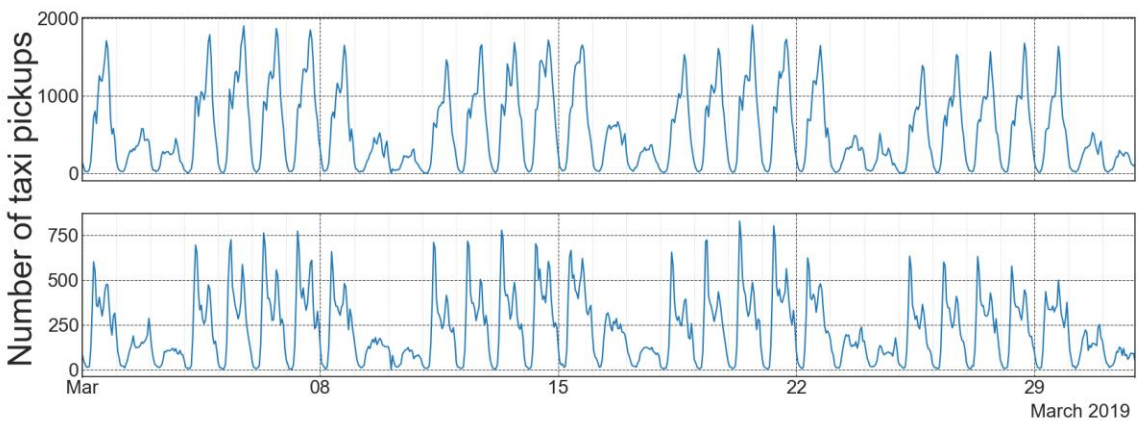
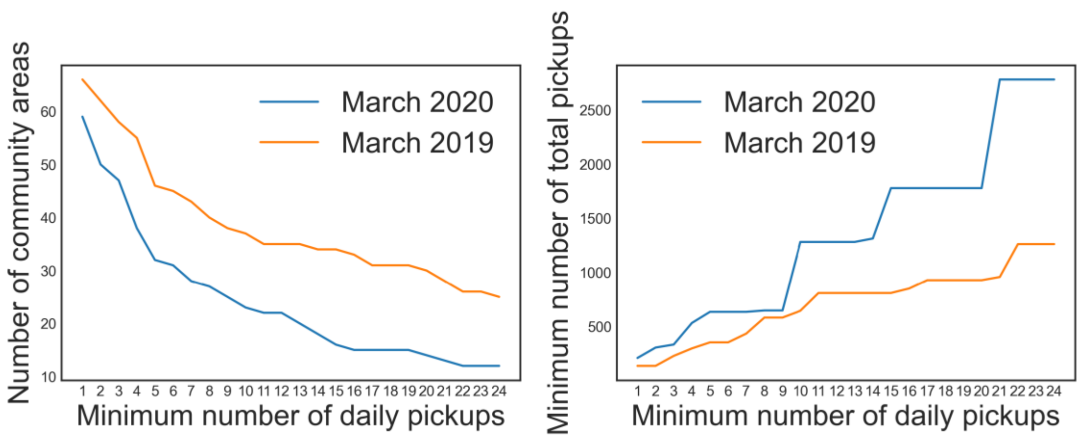


Figure 3. Examples of taxi pickup time series in two community areas in March 2019.

In addition, some community areas have very few records during the study period and do not have insufficient data for time series decomposition. Therefore, this research used a filter to remove community areas with very few daily records. The threshold of the filter is defined as the minimum number of taxi pickups in one day in each community area. When increasing the minimum number of daily taxi pickups, the number of community areas will decrease, and the minimum number of total pickups will increase. The relationship between the numbers of community areas and the minimum number of total pickups is used as a reference for the threshold setting.

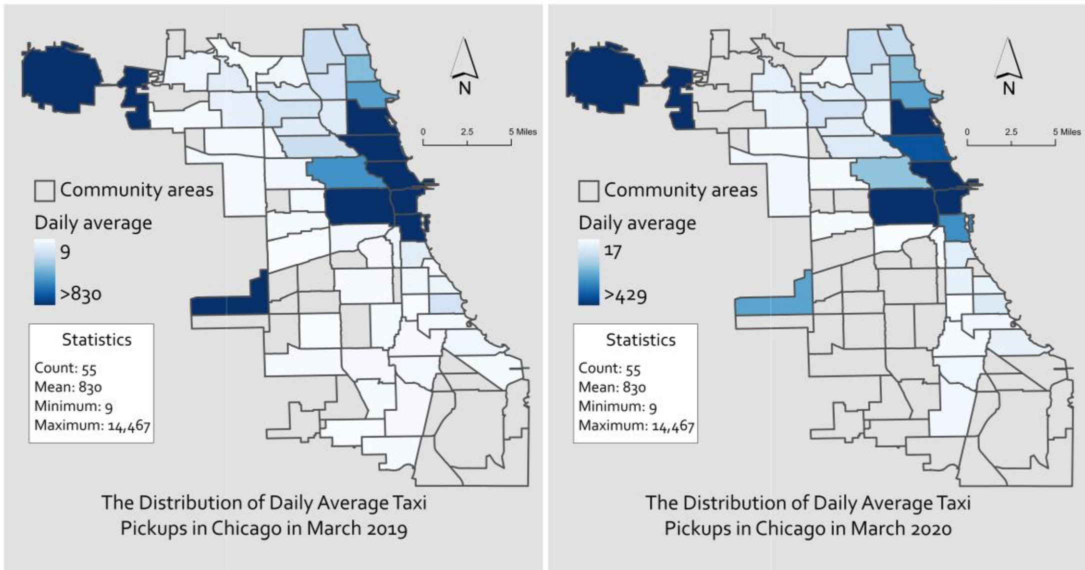


(a) **(b)**

Figure 4. The filtering results when changing the minimum number of daily pickups: (a) the number of community areas; (b) the minimum number of total pickups.

Figure 4 shows how the minimum number of daily pickups influences the number of community areas and the minimum number of total pickups among community areas. Using the elbow method, this research chose four as the cutoff value to keep comparable community areas and records in both periods. After data filtering, there are 55 community areas in March 2019 and 38 community areas in March 2020. Figures 5a and 5b visualize the quantitative distribution of daily average pickups in community areas after data filtering in March 2019 and March 2020. These two maps reflect a disproportional distribution of taxi pickups in different community areas. In both March

2019 and March 2020, the taxi pickups were concentrated in the central area with thousands of records, while other community areas only have several records.



(a) (b)

Figure 5. The quantitative distribution of taxi pickup numbers in community areas after the filter: (a) March 2019; (b) March 2020.

4.2 Time Series Decomposition

The second step is to apply the time series analysis method, and this research uses STL decomposition. The STL method is based on a non-parametric regression model, which is locally estimated scatterplot smoothing (LOESS) (Cleveland and Devlin 1988; Cleveland 1990; Zhu and Guo 2017). LOESS calculates the dependent variable using a k-nearest-neighborhood smoothing method that computes the linear combination of independent variables with different weights. The weight *W(u)* is generally calculated by a tricube weight function:

$$W(u) = \begin{cases} (1-u^3)^3 & \text{for } 0 \le u < 1 \\ 0 & \text{for } u \ge 1 \end{cases} \quad (3).$$

In details, the neighborhood weight $v_i(x)$ for any position x_i can be calculated as:

$$v_i(x) = W\left(\frac{|x_i - x|}{\lambda_n(x)}\right) \quad (4).$$

When introducing equation (4) to time series analysis, x is the time point for the current observation, $\lambda_n(x)$ is the distance of n temporal granularities from x, and x_i is any neighbor time point of x. The weight of x_i is determined by a distance ratio: the distance of the point x_i to current point x divided by the size of the smoothing window $\lambda_n(x)$. Therefore, the weight will decrease to zero when x_i gets away from x to the n^{th} neighbor time point. Generally, n is a parameter of LOESS that can be modified based on practical needs. By changing this parameter, STL can extract variation to components from the original time series. The main structure of STL contains three LOESS smoothers, and each smoother has an n parameter. Generally, these parameters have default values when the period parameter is assigned. The period is determined by the seasonal cycle. For example, for the taxi data used in this research, if aggregating the trip records by an hour, the period will be 24 as each day is a cycle of 24 hours. If aggregating the records by day, the period will be 7 as each week is a cycle of seven days. Therefore, the period parameter is set as 24 to extract daily seasonality. At the end of the time series decomposition, there will be series of trends and seasonality in each community area.

4.3 Clustering Analysis

The third step is analyzing the temporal patterns among community areas using clustering analysis. This research first measures the Euclidean distance of trend or seasonality series between different community areas. To investigate the similarity of time series in different community areas, we calculate the distance matrix based on Equation (6), which is used as the input of the clustering analysis. In the trend distance matrix T, each element t^{ab} represents the distance between each pair of trend series:

$$t^{ab} = \sqrt{\sum_{i=1}^{N}\left(t_i^a - t_i^b\right)^2} \qquad (5).$$

Equation (5) calculates the Euclidean distance t^{ab} between the trend series in community areas *a* and *b*, in which *i* is the index of the trend series, *N* is the length of the series, t_i^a and t_i^b are the observations at the time point *i* in the two series. Therefore, the trend distance matrix is a diagonal matrix containing the distance of trend in any two community areas. The seasonality distance matrix uses a calculation similar to the trend distance matrix. Then this research uses hierarchical clustering with an agglomerative method. The cluster number is determined by a rule of thumb and it can be set as $\sqrt{\frac{n}{2}}$, where *n* is the number of community areas (Han, Kamber, and Pei 2011). The clustering analysis focuses on grouping the community areas into clusters based on the trend and seasonality patterns extracted from the time series.

More importantly, this research assigns weights to trend and seasonality series to investigate their effects on the clustering results. The weighted series is calculated by this equation:

$$D = w \times T + (1 - w) \times S \qquad (6).$$

In equation (6), *D* is the weighted series, *T* is the trend series, *S* is the seasonality series, and *w* is a weighting parameter ranging from 0 to 1. When changing *w*, *D* will be a combination of *T* and *S* based on the weights assigned. With different weights assigned to trend and seasonality, the clustering analysis will get different results.

In addition, the taxi pickup patterns in different community areas are related to the urban land uses (Liu, et al. 2012b). Therefore, the next step in the discussed method is to find urban clusters and their connections with urban land uses. Figure 6 is the map of urban land uses in Chicago. The land uses are aggregated into five types: Residential, Commercial, Institutional, Industrial, and TRANS/COMM/UTIL/WASTE, which stands for transportation, communication, utilities, and waste facilities (Chicago Metropolitan Agency for Planning 2020). Each land use type has a specific social function and urban dynamics through different work schedules. The urban dynamics can be reflected in the transportation and thus the temporal patterns in the taxi pickups are closely connected to the urban land uses. Therefore, analyzing the connections between clusters of taxi pickup series and urban land uses can find the spatial patterns of taxi pickups in the urban area.

The community areas have mixed land uses in different proportions. For example, community areas 8, 32, and 33 are the Near North Side, the Loop, and the Near South Side, respectively. These three community areas are the central area of Chicago with most commercial places but also have some residential, institutional, and industrial land uses. To quantify the land use, this research adopted a method developed by Liu, et al. (2012b based on this equation:

$$D_i^c = \frac{L_i^{Avg\_c}}{L_i^{Avg\_all}} \quad (7).$$

In equation (7), the density (D_i^c) for each land use type i in cluster c, is calculated by dividing the average number of land use i in cluster c ($L_i^{Avg\_c}$) by the average number of that land use type in all community areas ($L_i^{Avg\_all}$).

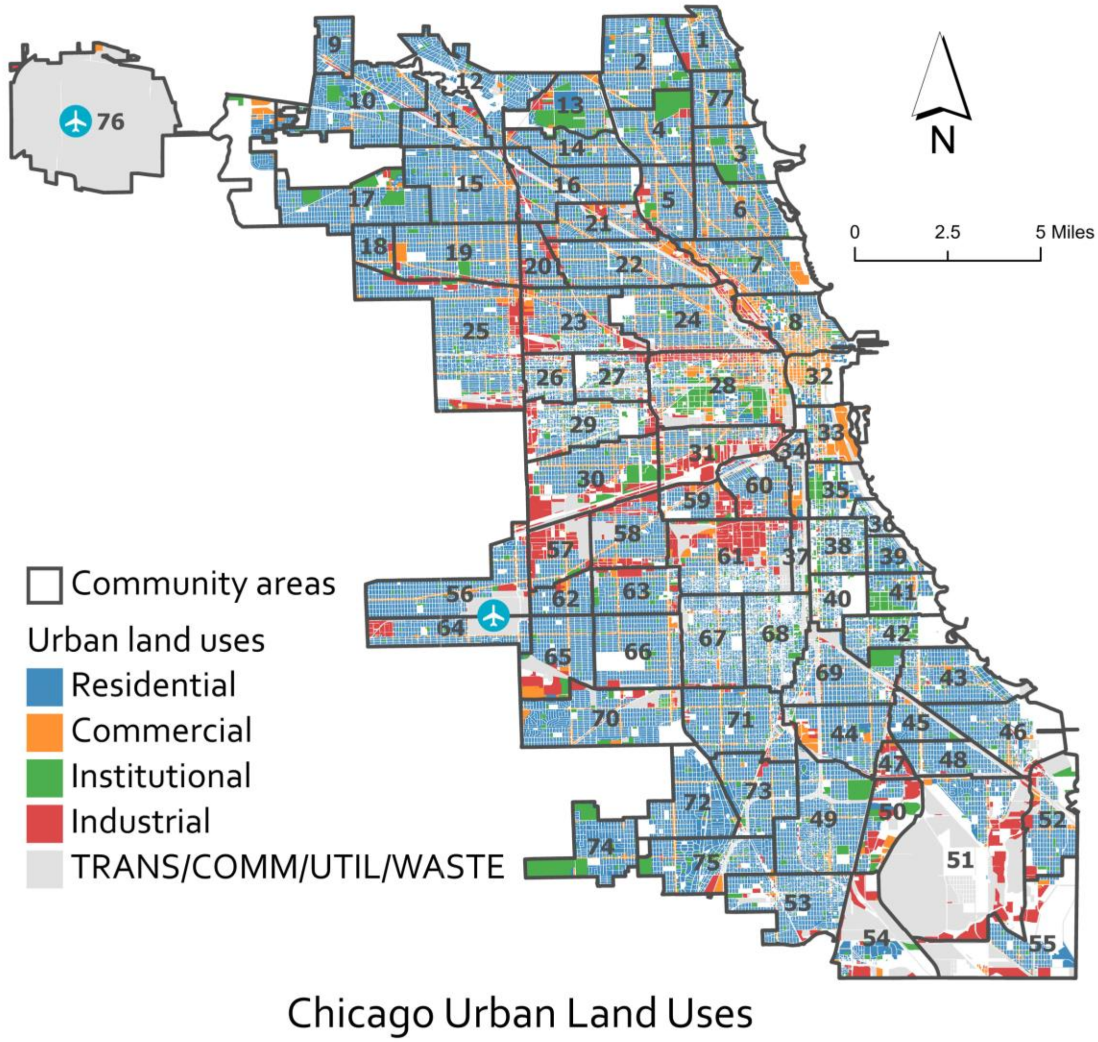


Figure 6. Map of Chicago urban land uses at the parcel level.

After extracting clustering results and the correspondence between clusters and land use patterns, this research compares the spatio-temporal patterns before and during COVID-19. The comparison is in three aspects: the differences of clustering results and patterns in these clusters, the differences in the cluster constitution change when adding the trend component, and the differences of the connections with urban land uses. The differences in the clustering results can indicate the total impacts of COVID-19 on taxi pickups in a city-wide range. In addition, comparing the difference of how the cluster constitutions and patterns change when increasing the trend ratio helps to investigate the trend and seasonality effects in these urban areas before and during COVID-19. Furthermore, to better understand the spatial differences of the clustering results, the connections with urban land uses can provide assumptions and interpretations for the patterns in clusters. The following section explains the results in detail.

5 RESULTS

5.1 Clustering Results with Different Trend/Seasonality Ratio in March 2019

Because the variability of the trend and seasonality not only exists in time but also in space, the clustering analysis can group community areas based on similar taxi trends or seasonality. Figure 7 and Figure 8 show clusters of seasonality, trend, and the weighted series of taxi pickups in community areas in March 2019. The weighted series is the addition of quartiles of trend and seasonality, so the clustering analysis is performed on the series of seasonality, weighted series (0.25trend + 0.75seasonality, 0.5trend + 0.5seasonality, 0.75trend + 0.25seasonality), and trend separately. Changing the trend/seasonality ratio in the weighted series helps us investigate how the short-term and long-term patterns will change the clustering results. For the hierarchical clustering analysis parameters, first, the cluster number is set to 5 as there are 55 community areas. The group method is agglomerative clustering and uses variance as the linkage criterion. Therefore, the 5 clusters are grouped by minimized variance.

5.1.1 Seasonality

Figure 7 shows the clustering results of taxi pickup seasonality during the month in the first column and the zoom-in views of the first two days in the second column. In the whole month view, the major ticks on the x-axis indicate Fridays, dividing the month into weeks. In the zoom-in view, the major ticks divide one day into four parts: early morning (12 am – 6 am), morning (6 am – 12 pm), afternoon (12 pm – 6 pm), and evening (6 pm – 12 am). Each part contains 6 hours. In each cluster, the daily fluctuations are almost the same during the month, and the red line represents the average daily fluctuations. Although the average series peaks are different among clusters, the bottoms are all in the early morning.

Although the average lines in cluster 0 and cluster 1 have peaks in the morning, these peaks do not accord with the series of community areas. The series in these community areas have aggregated patterns with uncertainty. The uncertainty here means high ingroup variance so the average line cannot represent the general patterns. While in other clusters, the patterns are specific with daily peaks in the morning, afternoon, or evening.

The unique patterns of taxi pickup seasonality in these clusters reflect different daily transportation behaviors among community areas. The patterns of series in cluster 0 and cluster 1 are uncertain with a high variance. Thus, the series are grouped together more likely by the fluctuation range instead of distinguishable shapes. The fluctuation range is small in clusters 0 and 1 because daily taxi pickup records in these community areas are insufficient to form a regular and distinctive temporal distribution. Cluster 2 and cluster 4 show a high taxi demand pattern in the afternoon and the evening. The close peak time indicates connections or similarities in urban functions of the community areas in these two clusters, like commercial, recreational, and working places. People in these places have commuting demands in the afternoon and the evening. However, cluster 3 shows high taxi demand in the morning in these community areas. People in these areas have a higher demand to commute to work, so there might be more residential places or urban transfer stations. From the different times of taxi demand in a day, seasonality can reflect the dynamics of daily schedules in different urban areas.

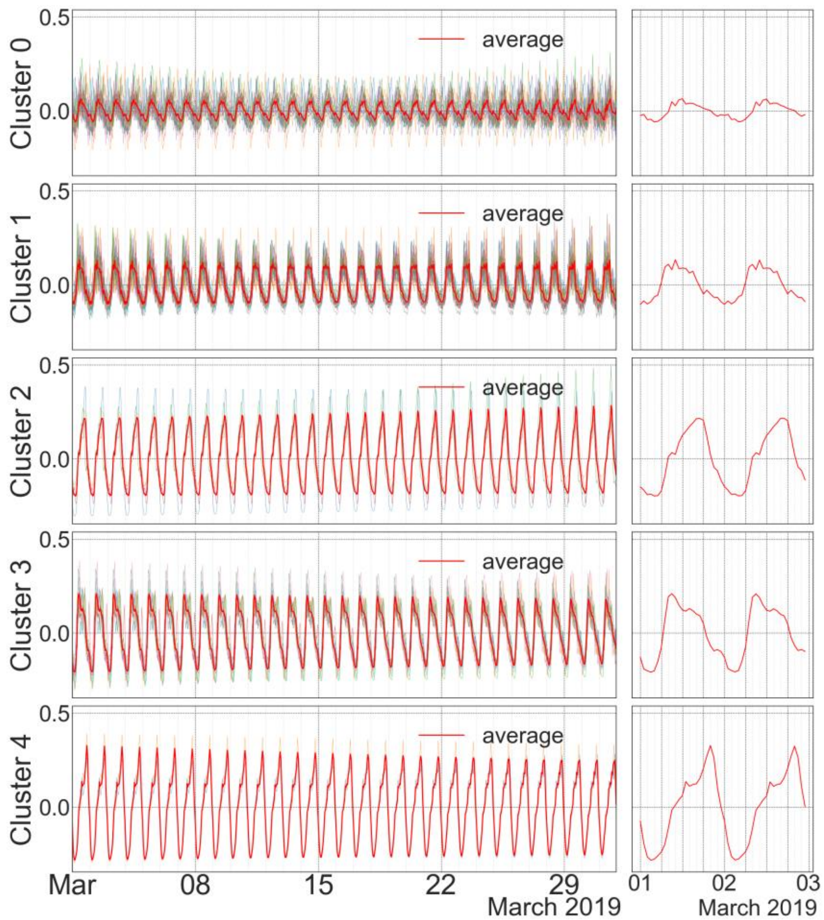


Figure 7. Clustering results of taxi pickup seasonality in March 2019.

5.1.2 Adding the trend component to the clustering analysis

When increasing the weight of trend from 0.25, 0.5, 0.75, to 1 in the weighted series, the clustering results changed accordingly. During this process, some clusters will disassemble, and the community areas will merge with other clusters. To find the connections between clusters on different weighted series, the cluster ID is assigned depending on the constitution of community areas in that cluster. If the cluster contains more than 50% community areas from the previous clustering results, it will keep the same ID as the previous cluster. If no more than 50% of community areas are the same as in the previous clusters, the cluster containing these community areas will have a new ID.

Figure 8 shows the clustering results of weighted series with different trend/seasonality weights. Figure 8a provides the general changes when increasing the trend ratio from 0.25, 0.5, to 0.75. During this process, some clusters dissolved such as clusters 2 to 5, while other clusters appeared and persist until the end like clusters 6 to 9. These clusters keep patterns of both daily and weekly fluctuations with different proportions. The weekly fluctuations are more substantial in some clusters like clusters 6, 8, and 9 when increasing the weight of the trend component. This fluctuation intensity shows a significant difference in taxi passenger flow on weekdays and weekends, which is likely caused by different social activities on different days in a week in these clusters. In other clusters where the weekly fluctuations do not look obvious, the trend component helps to differentiate the series with slightly different weekly patterns from these clusters such as cluster 0 and cluster 7. When increasing the trend ratio, the daily fluctuation became smoother, and the weekly patterns are more apparent in each cluster. In the end, when the trend ratio is 1, the clustering results are all based on the trend component, and the clusters are shown in Figure 8b. It can be seen from Figure 8b that most clusters have unique weekly patterns while cluster 0 still contains series with uncertain patterns.

The daily and weekly patterns in these clusters can indicate urban dynamics and social functions in these areas. For example, the weekly drop on weekends and daily increase in the morning may indicate more workplaces in the community areas in cluster 6. The increase of taxi demand on Fridays and weekends in cluster 9 indicates more social activities on weekends in these community areas. In addition, the large increase of taxi pickups on March 16th and 17th in cluster 9 might be related to the celebration of St. Patrick's Day. On the contrary, the uncertain patterns in some clusters are probably due to insufficient records to form a regular temporal distribution, or the randomness of the temporal distribution of taxi pickups in these areas, which is an inevitable component of human mobility (Song et al. 2010).

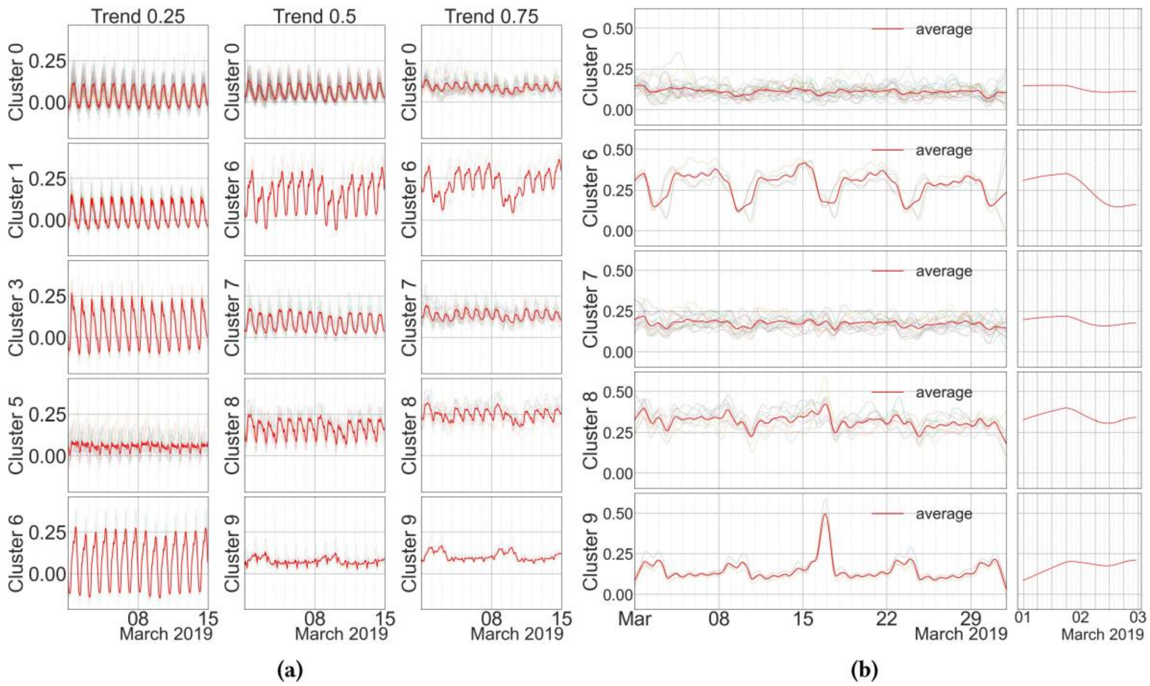


Figure 8. Clustering results of taxi pickups in March 2019 when adding the trend component: (a) partial views of results in the first two weeks in March when trend/seasonality ratios are 0.25: 0.75, 0.5: 0.5, and 0.75: 0.25; (b) results when trend/seasonality is 1: 0.

The effects of the trend and seasonality components are revealed in the transition from the daily patterns to the weekly patterns. When adding the trend ratio, the series in the clustering results contain fewer daily fluctuations but more weekly fluctuations. This causes the cluster constitutions and short/long-term patterns to change. These patterns can help identify the urban function in the clusters of community areas where the daily patterns are uncertain, but if the weekly patterns are still uncertain, it might indicate low taxi demand or randomly temporal distribution of taxi pickups in those areas.

Changing the trend and seasonality ratio can affect the clustering results on the community area constitutions and the clusters' aggregated patterns. The constitutions and patterns change gradually along with the ratio change. Investigating the connections between clusters of different ratios can identify how the combination of seasonality and trend affects the clustering analysis. Table 1 displays the existing range and patterns for each cluster. The existing range refers to when the cluster exists in the process of increasing the trend ratio. The patterns include the daily peaks or weekly fluctuation if applicable. Clusters appear and disappear in the increasing trend ratio process, with daily patterns or weekly patterns, or neither patterns are apparent.

Clusters from 0 to 5 appear when the trend ratio is 0 or 0.25, but only clusters 2, 3, 4, and 6 have daily patterns. Cluster 5 is more likely a transitional cluster with high uncertainty in the patterns. Cluster 1 has no apparent patterns either, although it exists during the whole process. When increasing the trend ratio to 0.5, cluster 6 to cluster 9 appear and exist to the end. These clusters all have weekly patterns, and cluster 6 has both daily and weekly patterns.

Table 1. Summary of taxi pickup patterns of clusters in March 2019

Cluster	Existing Range	Patterns
0	0 to 1	Uncertain
1	0 to 0.25	Uncertain
2	0	Afternoon peak
3	0 to 0.25	Morning peak
4	0	Evening peak
5	0.25	Highly uncertain
6	0.25 to 1	Evening peak, Weekends drop
7	0.5 to 1	Weekends slight drop
8	0.5 to 1	Sunday drop
9	0.5 to 1	Friday and weekends increase

5.1.3 The relationship of urban land uses and clustering results in March 2019

After extracting different pickup patterns in the clusters, it helps to connect these patterns with urban land uses to help us better understand how and why these patterns are distributed in the clusters. In March 2019, the clustering results have major changes when increasing the trend ratio to 0.25 and 0.5, and accordingly, the spatial distribution of the clusters and the land use constitutions mainly changed before the 0.5 trend ratio. To visualize the spatial distributions and land uses of these mainly changed clusters, Figure 9 contains the distribution maps and land use constitution charts at the trend ratios of 0, 0.25, and 1. Different constitutions of land use density among the clusters can reflect urban functions in these areas. The urban functions determine the social activities, influencing taxi passenger flow and pickup patterns in an urban area. Although, the spatial resolution of the community area is much coarser than the resolution of land use data, there are obvious relationships between land use constitutions and patterns of taxi pickups. These patterns, especially with specific daily fluctuations when the trend ratio is 0 or 0.25, are more likely in the clusters with higher commercial and institutional land uses or the areas with special urban functions than uncertain patterns.

For example, as Figures 9a and 9b show, clusters 2, 3, and 6 have higher commercial and institutional land use density than the average, and cluster 4 has airports in its two community areas. In these clusters, the daily fluctuations are prominent, which can reflect connections between clustering results and urban land uses. For clusters 0, 1, and 5, the daily patterns tend to be uncertain as they have fewer commercial or institutional land uses. When increasing the trend ratio to more than 0.5, the patterns of the clusters might be also related to the urban functions with compositive land uses. For example, cluster 9 has a higher land use density of all four types and is the only cluster with taxi pickup increases on weekends. This increase is probably due to the integrated and compositive land uses and more social activities on weekends. In addition, when increasing the trend ratio in the clustering analysis, some community areas containing a higher density of compositive land uses will be grouped together, and the taxi pickup fluctuations will show weekly or long-term patterns.

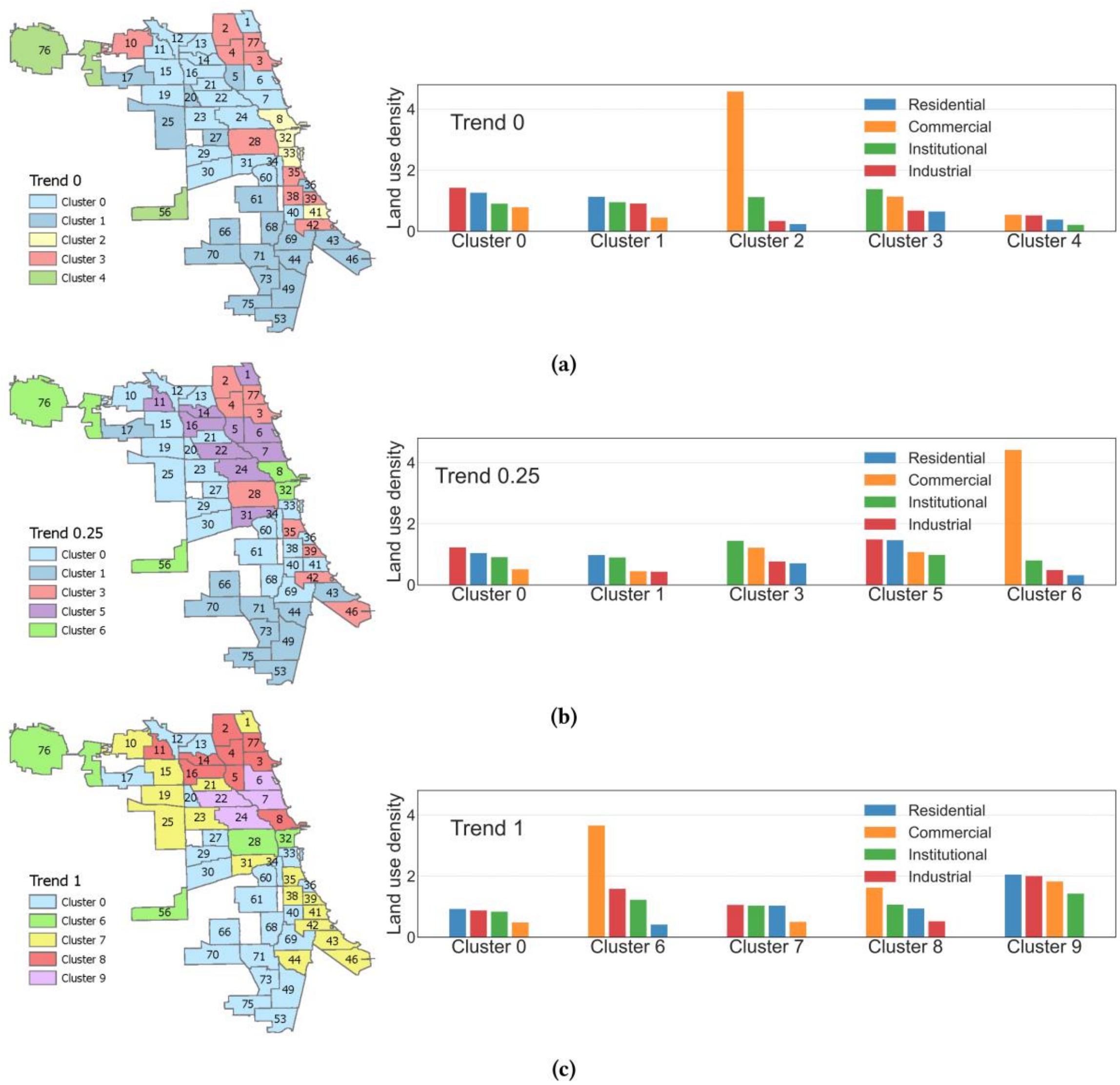


Figure 9. Map and land use density of community area clusters in March 2019 with trend/seasonality ratios: (a) 0: 1; (b) 0.25: 0.75; (c) 1: 0.

5.2 Comparing the patterns between March 2019 and March 2020

This section compared the spatio-temporal patterns of clusters in March 2019 and March 2020 (before and during COVID-19). The differences and implications are in three aspects. First, the differences in clustering results and patterns in these clusters can reflect the total impacts of COVID-19 on taxi pickups in a city-wide range. Secondly, when adding the trend component, the cluster constitution changes indicate that the trend and seasonality effects in these urban areas are different before and during COVID-19. Additionally, the connections between urban land uses and the clustering results also support the different patterns in COVID-19.

5.2.1 Clustering results and patterns in March 2020

Compared to March 2019, the taxi record numbers sharply decreased in March 2020, and there are only 38 community areas with at least four records per day. The clustering analysis on taxi pickup numbers also uses the weighted series by increasing the trend ratio from 0, 0.25, 0.5, 0.75 to 1, and the cluster number is set to 4 according to 38 community areas. Figures 10a, 10b, and 10c are the clustering results of seasonality series, weighted series of 0.5trend + 0.5seasonality, and trend series. When increasing the trend ratio, the cluster constitutions in March 2020 change less, compared to the cluster constitutions in March 2019. There is only one cluster ID change in March 2020. Another significant difference is the abnormal patterns after mid-March 2020.

Before mid-March, the clustering results have daily and weekly patterns from trend and seasonality shown in Figures 10a and 10c. The daily fluctuations have diverse distributions during the diurnal period, indicating regular daily dynamics of taxi demand. The weekly fluctuations are all decreasing on weekends, indicating less human mobility and social activities in these areas than that in March 2019.

The dramatic decrease of taxi pickups in all clusters appears after mid-March 2020. Both daily and weekly fluctuations decrease. Although there had been COVID-19 cases reported in Chicago since January and taxi pickups reduce on weekends, the largest impact on taxi demand during the early period of COVID-19 is more related to government responses as they initiated closing public places and stay-at-home orders in mid-March 2020.

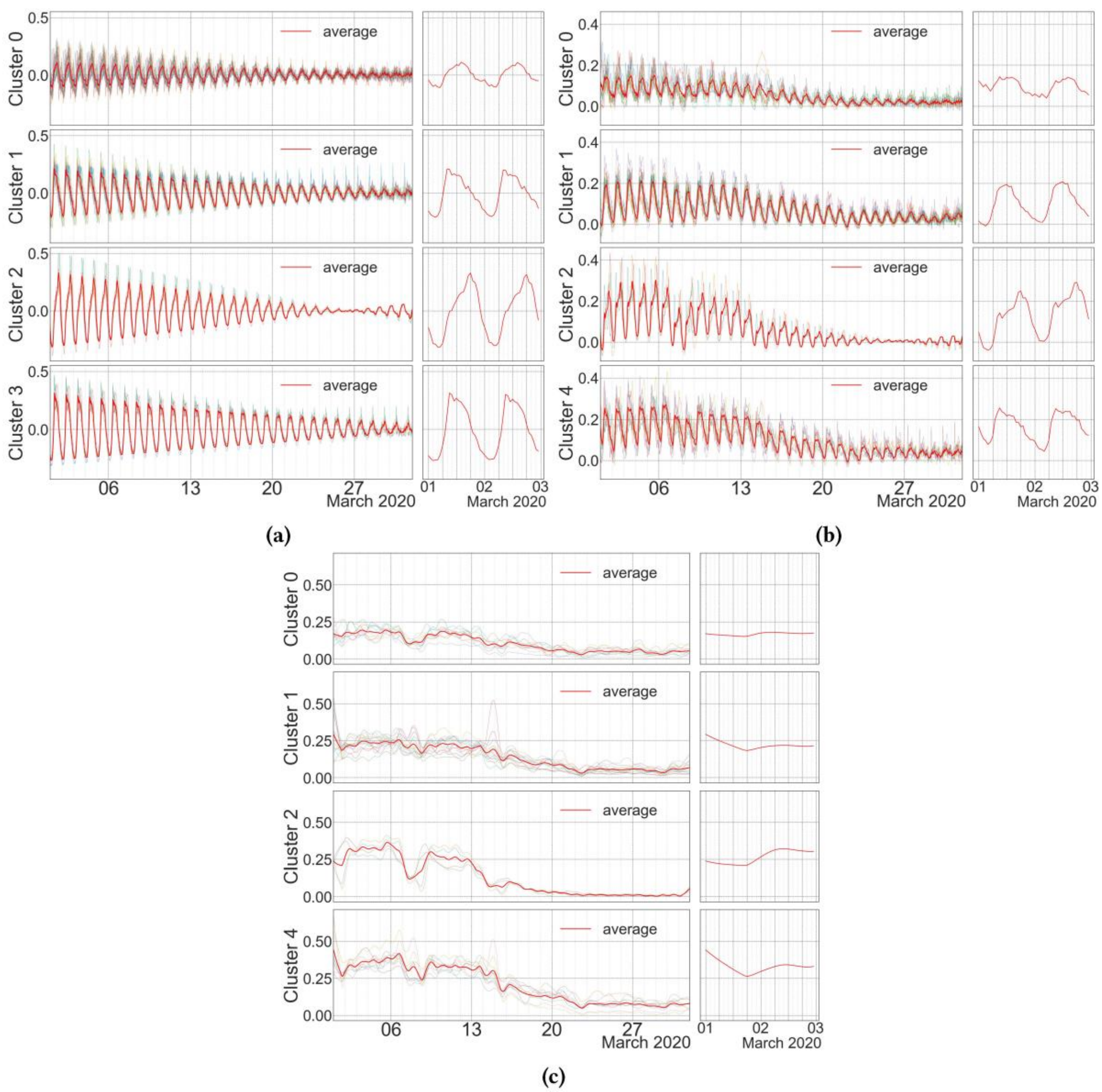


Figure 10. Time series in clusters March 2020 with trend/seasonality ratios: (a) 0: 1; (b) 0.5: 0.5; (c) 1: 0.

Compared to the clusters in March 2019, the clustering results in March 2020 have less change in the constitutions of clusters when increasing the trend ratio. Table 2 shows the summary of taxi pickup patterns in clusters in early March 2020. Most clusters exist during the whole period, including clusters 0, 1, and 2. Cluster 3 disappears and cluster 4 forms when the trend ratio is more than 0.5. The four cluster IDs indicate that the cluster constitutions remain stable when changing the trend ratio, so the clusters are not influenced much by daily or weekly patterns in the seasonality or trend components. In addition, the daily and weekly patterns are more unified than those in March 2019. The daily patterns are distinguishable only in cluster 2 and cluster 3. The weekly patterns show a decrease in taxi pickups on weekends, but each cluster shows different fluctuation ranges. After comparing the patterns, the connections and changes among clusters are discussed in the next subsection.

Table 2. Summary of daily and weekly patterns of taxi pickup numbers in early-March 2020

Cluster	Existing Range	Patterns
0	0 to 1	Weekends slight drop
1	0 to 1	Weekends slight drop
2	0 to 1	Evening peak, Weekends drop
3	0 to 0.25	Morning peak
4	0.5 to 1	Weekends drop

5.2.2 How clusters change in 2019 and 2020 when assigning different weights for trend and seasonality

Figure 11 displays alluvial diagrams showing how clusters change with increasing trend ratio in March 2019 and March 2020. The alluvial diagram was used to visualize the urban cluster interchange over time in the community detection studies (Zhong et al.

2014; Zhong et al. 2015). Rather than showing the community changes over time, here the diagrams present how the clustering results change with different trend and seasonality ratios. From the left to the right of the diagram, the clustering results have changes in the structure and cluster constitution when adding the trend component. The stages of trend ratios are represented by columns. Each column of the diagram contains clusters represented by bars. The bar length is proportionate to the number of community areas in that cluster, and the label T[number]_[number] represents Trend[ratio]_[cluster ID].

Figure 11 shows that the changes of clusters in March 2019 are more substantial than in March 2020 when increasing the trend ratio to 0.25 and 0.5. Figure 11a shows that when adding the trend component, the community area exchange mostly occurs from the cluster with uncertain patterns such as cluster 0 and cluster 1. These community areas merge to a new cluster with more specific patterns such as clusters 7 to 9 if the trend component helps them reduce the uncertainty in the patterns. On the contrary, if the trend patterns are still uncertain, these community areas will remain in cluster 0.

Figure 11b shows that, according to the less diverse patterns of taxi pickups, the cluster constitutions in March 2020 are more stable than those in March 2019. Although the major exchange of community areas also occurs when increasing the trend ratio from 0 to 0.5, the main structure of the clustering results does not change much. Most clusters remain in the results when adding the trend component, except cluster 3 replaced by cluster 4. In addition, cluster 4 consists of community areas from the other four clusters. This is probably because the trend in these clusters has similar patterns with each other as the taxi pickup numbers all decrease on weekends.

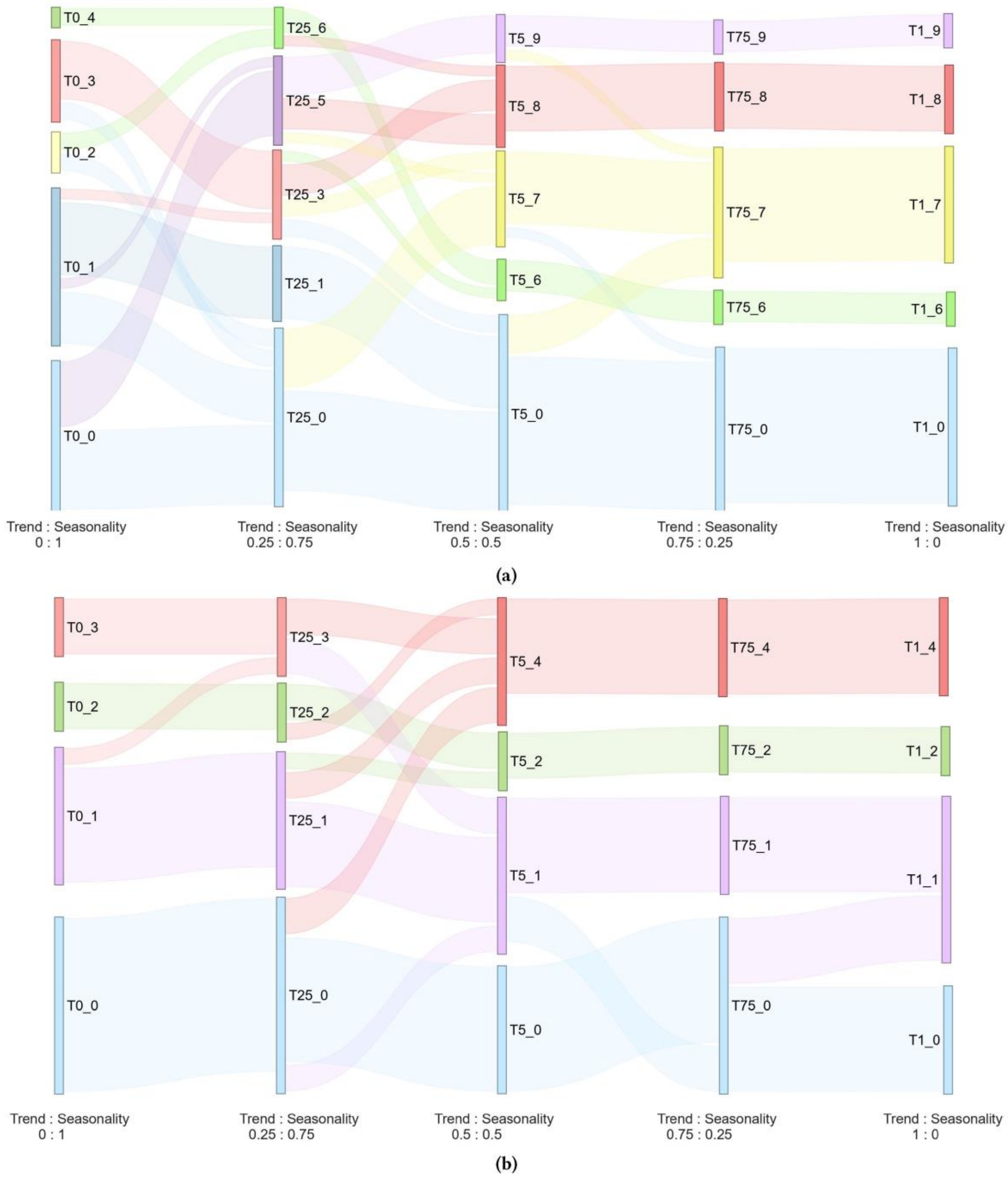


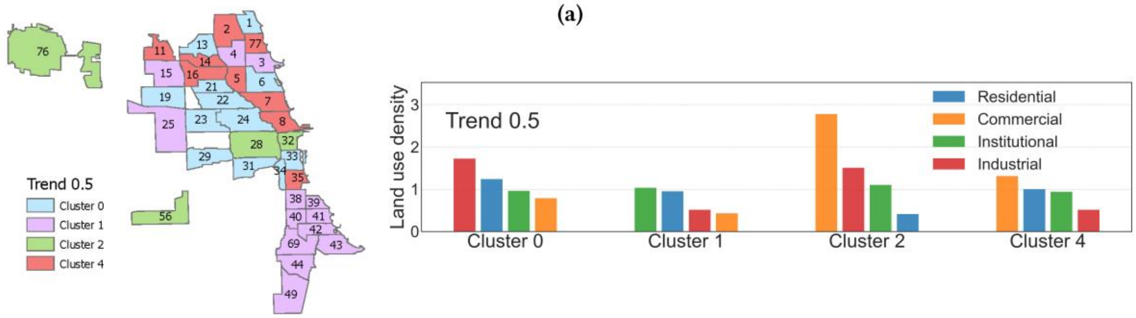
Figure 11. Cluster constitution changing flow with different ratios of trend and seasonality in (a) March 2019; (b) March 2020.

5.2.3 The relationship of urban land uses and clustering results in March 2020

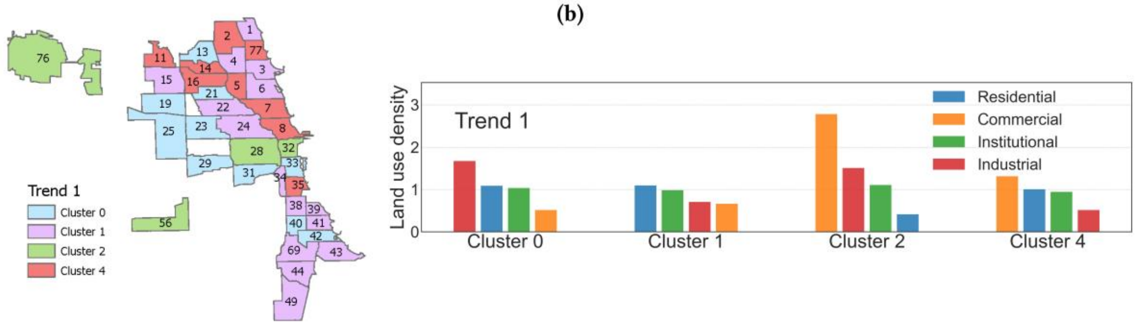
Figure 12 shows the spatial distribution of clustering results and urban land use constitution in these clusters in March 2020. Compared with the land use patterns in March 2019, the constitutions of community areas in the clusters do not change much when increasing the trend ratio, and the land use constitutions also have fewer changes in March 2020. Although these clusters have unique urban land use constitutions and consistent taxi pickup patterns, the stable urban land use constitutions indicate fewer effects in the transition from seasonality to the trend in March 2020. This is probably because the clusters are grouped more based on the fluctuation range instead of specific temporal distributions. For example, over the weekly period, the patterns in these clusters have less diversity in the temporal distribution but more differences in the fluctuation range. These results indicate that COVID-19 has affected the taxi trips in Chicago by reducing the diversity of taxi pickup temporal distributions and influencing the fluctuation range in different areas.

Due to the differences between 2019 and 2020 patterns, it seems that changing the weights for trend/seasonality does not have that much impact on patterns in 2020. This indicates when the daily/weekly fluctuations are dominant in the general fluctuation of the time series during the research period, extracting the trend/seasonality components is the most effective method to investigate the long/short-term patterns. Both trend and seasonality can impact the results of clustering, but the range of fluctuation seems to mask the daily/weekly patterns in March 2020, so the clusters did not change much with different trend/seasonality ratios. As the fluctuation range is stable in the same week, we can consider modifying the normalization process by normalizing the pickups in a week instead of a month to reduce the effect from fluctuation range in the whole research period, which can be an alternative method in the future research.

(a)



(b)



(c)

Figure 12. Map and land use density of community area clusters in March 2020 with different trend/seasonality ratios: (a) 0: 1; (b) 0.5: 0.5; (c) 1: 0.

6 CONCLUSION

This research used time series decomposition to analyze the temporal patterns in the trend and seasonality from taxi trip records in the Chicago community areas. The trend and seasonality of taxi pickups contain weekly and daily fluctuations as long-term and short-term patterns. To further analyze the spatial patterns of these time series, this research implemented clustering analysis on the trend, seasonality, and their weighted series, and discussed the implications of urban land uses in these clusters. The findings and contributions of this research include:

1. effects of seasonality and trend components in the urban clusters and implications from urban land uses. Assigning different weights of seasonality the trend can help us investigate various daily and weekly patterns among community areas and distinguish community areas with uncertain patterns. The implications from urban land uses indicate the clusters with apparent daily patterns of taxi pickups have more commercial or institutional land uses or have special areas such as airports. The clusters with distinctive weekly patterns might relate to compositive urban land uses for supporting more social activities and increasing taxi passenger flow;
2. empirical findings when comparing the taxi data between March 2019 and March 2020. In March 2020, the major decrease in taxi trip number is probably brought by the government responses to COVID-19 after mid-March. Besides the number decrease, the impact of COVID-19 on human mobility reflects on the lack of diversity in the taxi trip temporal distribution among community areas. The daily/weekly patterns were masked by the effect of the fluctuation range in the research period of March 2020. These findings

help us get a better understanding of how COVID-19 affected taxi trips as human mobility and have a better consideration in designing future research to effectively implement the method we developed in this research.

In conclusion, this research thoroughly discussed how to use the trend and seasonality from time series decomposition to extract temporal patterns in the taxi trips in Chicago and compare the patterns before and during COVID-19 to investigate the impacts of COVID-19 on human mobility. There are some limitations in this research. First, the spatial resolution is at the community areas and is too coarse to reflect urban heterogeneity within a community area. Second, this research uses taxi trips, which can only represent part of the population. Further work can focus on higher spatial resolution data or floating car data with a more diverse constitution. In addition, the methodology developed in this research can be applied to other datasets or other periods of Chicago taxi data to find more interesting patterns in human mobility.

REFERENCES

Andris, Clio. 2016. "Integrating Social Network Data into Gisystems." *International Journal of Geographical Information Science* 30, no. 10: 2009-2031.

Bajardi, Paolo, Chiara Poletto, Jose J Ramasco, Michele Tizzoni, Vittoria Colizza, and Alessandro Vespignani. 2011. "Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1n1 Pandemic." *PloS one* 6, no. 1: e16591. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3031602/pdf/pone.0016591.pdf.

Belik, Vitaly, Theo Geisel, and Dirk Brockmann. 2011. "Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases." *Physical Review X* 1, no. 1: 011001.

CDC. 2020a. "How Covid-19 Spreads." Coronavirus Disease 2019 (COVID-19). Last modified 6/16/2020. Accessed 8/13, 2020. Form of Item. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html.

---. 2020b. "Protect Yourself When Using Transportation." Coronavirus Disease 2019 (COVID-19). Last modified 8/3/2020. Accessed 8/13, 2020. Form of Item. https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/using-transportation.html#RideShare.

Chen, Cynthia, Jingtao Ma, Yusak Susilo, Yu Liu, and Menglin Wang. 2016. "The Promises of Big Data and Small Data for Travel Behavior (Aka Human Mobility) Analysis." *Transportation research part C: emerging technologies* 68: 285-299.

Chicago Data Portal. 2016. "Chicago Taxi Data Released." Last modified November 16, 2016. Form of Item. https://digital.chicago.gov/index.php/chicago-taxi-data-released/.

Chicago Metropolitan Agency for Planning. 2020. *Land Use Inventory for Northeast Illinois, 2015*. 11/23/2020 ed. CMAP Data Hub.

Chicago Tribune. 2020. "Covid-19 in Illinois, the U.S. And the World: Timeline of the Outbreak." Last modified MAY 15, 2020. Form of Item. https://www.chicagotribune.com/coronavirus/ct-viz-coronavirus-timeline-20200507-uvrzs32nljabrpn6vkzq7m2fpq-story.html.

City of Chicago. 2020. "City of Chicago :: Covid Dashboard." City of Chicago. Form of Item. https://www.chicago.gov/city/en/sites/covid-19/home/covid-dashboard.html.

Cleveland, Robert B. 1990. "Stl: A Seasonal-Trend Decomposition Procedure Based on Loess. 1990." *Journal of Official Statistics* 6, no. 1: 3-73.

Cleveland, William S, and Susan J Devlin. 1988. "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American statistical association* 83, no. 403: 596-610.

Croce, Antonello Ignazio, Giuseppe Musolino, Corrado Rindone, and Antonino Vitetta. 2019. "Transport System Models and Big Data: Zoning and Graph Building with Traditional Surveys, Fcd and Gis." *ISPRS International Journal of Geo-Information* 8, no. 4: 187.

De Fabritiis, Corrado, Roberto Ragona, and Gaetano Valenti. 2008. *Traffic Estimation and Prediction Based on Real Time Floating Car Data*. *2008 11th international IEEE conference on intelligent transportation systems*: IEEE.

Deng, Cheng. 2014. "Time Series Decomposition Using Singular Spectrum Analysis."

Eagle, Nathan, and Alex Sandy Pentland. 2006. "Reality Mining: Sensing Complex Social Systems." *Personal and ubiquitous computing* 10, no. 4: 255-268.

Fang, Zhiming, Zhongyi Huang, Xiaolian Li, Jun Zhang, Wei Lv, Lei Zhuang, Xingpeng Xu, and Nan Huang. 2020. "How Many Infections of Covid-19 There Will Be in the" Diamond Princess"-Predicted by a Virus Transmission Model Based on the Simulation of Crowd Flow." *arXiv preprint arXiv:2002.10616*.

Gao, Mengdan, Tongyu Zhu, Xuejin Wan, and Qi Wang. 2013. *Analysis of Travel Time Patterns in Urban Using Taxi Gps Data*. *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*: IEEE.

Gao, Song, Jinmeng Rao, Yuhao Kang, Yunlei Liang, and Jake Kruse. 2020. "Mapping County-Level Mobility Pattern Changes in the United States in Response to Covid-19." *SIGSPATIAL Special* 12, no. 1: 16-26.

Goddard, John Burgess. 1970. "Functional Regions within the City Centre: A Study by Factor Analysis of Taxi Flows in Central London." *Transactions of the Institute of British Geographers*: 161-182.

Gonzalez, Marta C, Cesar A Hidalgo, and Albert-Laszlo Barabasi. 2008. "Understanding Individual Human Mobility Patterns." *nature* 453, no. 7196: 779-782. https://www.nature.com/articles/nature06958.

National Bureau of Economic Research, 2020. *Tracking Public and Private Response to the Covid-19 Epidemic: Evidence from State and Local Government Actions.* by Gupta, Sumedha, Thuy D Nguyen, Felipe Lozano Rojas, Shyam Raman, Byungkyu Lee, Ana Bento, Kosali I Simon, and Coady Wing.

Haffajee, Rebecca L, and Michelle M Mello. 2020. "Thinking Globally, Acting Locally—the Us Response to Covid-19." *New England Journal of Medicine* 382, no. 22: e75. https://www.nejm.org/doi/pdf/10.1056/NEJMp2006740?articleTools=true.

Han, Jiawei, Micheline Kamber, and Jian Pei. 2011. "Data Mining Concepts and Techniques Third Edition." *The Morgan Kaufmann Series in Data Management Systems* 5, no. 4: 83-124.

Hyndman, Rob J, and George Athanasopoulos. 2018. *Forecasting: Principles and Practice*: OTexts.

Jiang, Bin, Junjun Yin, and Sijian Zhao. 2009. "Characterizing the Human Mobility Pattern in a Large Street Network." *Physical Review E* 80, no. 2: 021136.

Kang, Chaogui, Xiujun Ma, Daoqin Tong, and Yu Liu. 2012. "Intra-Urban Human Mobility Patterns: An Urban Morphology Perspective." *Physica A: Statistical Mechanics and its Applications* 391, no. 4: 1702-1717.

Koopmans, Lambert H. 1995. *The Spectral Analysis of Time Series*: Elsevier.

Kraemer, Moritz UG, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Louis Du Plessis, Nuno R Faria, Ruoran Li, and William P Hanage. 2020. "The Effect of Human Mobility and Control Measures on the Covid-19 Epidemic in China." *Science* 368, no. 6490: 493-497. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7146642/pdf/abb4218.pdf.

Liu, Xi, Li Gong, Yongxi Gong, and Yu Liu. 2015. "Revealing Travel Patterns and City Structure with Taxi Trip Data." *Journal of Transport Geography* 43: 78-90.

Liu, Yu, Chaogui Kang, Song Gao, Yu Xiao, and Yuan Tian. 2012a. "Understanding Intra-Urban Trip Patterns from Taxi Trajectory Data." *Journal of geographical systems* 14, no. 4: 463-483.

Liu, Yu, Xi Liu, Song Gao, Li Gong, Chaogui Kang, Ye Zhi, Guanghua Chi, and Li Shi. 2015. "Social Sensing: A New Approach to Understanding Our Socioeconomic Environments." *Annals of the Association of American Geographers* 105, no. 3: 512-530.

Liu, Yu, Fahui Wang, Yu Xiao, and Song Gao. 2012b. "Urban Land Uses and Traffic 'Source-Sink Areas': Evidence from Gps-Enabled Taxi Data in Shanghai." *Landscape and Urban Planning* 106, no. 1: 73-87.

Mazimpaka, Jean Damascène, and Sabine Timpf. 2016. "Trajectory Data Mining: A Review of Methods and Applications." *Journal of Spatial Information Science* 2016, no. 13: 61-99.

Moreira-Matias, Luis, Joao Gama, Michel Ferreira, Joao Mendes-Moreira, and Luis Damas. 2013. "Predicting Taxi–Passenger Demand Using Streaming Data." *IEEE Transactions on Intelligent Transportation Systems* 14, no. 3: 1393-1402.

Oldekop, Johan A, Rory Horner, David Hulme, Roshan Adhikari, Bina Agarwal, Matthew Alford, Oliver Bakewell, Nicola Banks, Stephanie Barrientos, and Tanja Bastia. 2020. "Covid-19 and the Case for Global Development." *World Development*: 105044.

Oliver, Nuria, Bruno Lepri, Harald Sterly, Renaud Lambiotte, Sébastien Deletaille, Marco De Nadai, Emmanuel Letouzé, Albert Ali Salah, Richard Benjamins, and Ciro Cattuto. 2020. *Mobile Phone Data for Informing Public Health Actions across the Covid-19 Pandemic Life Cycle*: American Association for the Advancement of Science.

Pan, Xiao-Ben. 2020. "Application of Personal-Oriented Digital Technology in Preventing Transmission of Covid-19, China." *Irish Journal of Medical Science (1971-)*: 1-2.

Qi, Guande, Xiaolong Li, Shijian Li, Gang Pan, Zonghui Wang, and Daqing Zhang. 2011. *Measuring Social Functions of City Regions from Large-Scale Taxi Behaviors*. *2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*: IEEE.

Rodrigue, Jean-Paul. 2016. *The Geography of Transport Systems*: Taylor & Francis.

Rodrigues, Filipe, Ioulia Markou, and Francisco C Pereira. 2019. "Combining Time-Series and Textual Data for Taxi Demand Prediction in Event Areas: A Deep Learning Approach." *Information Fusion* 49: 120-129.

Shumway, Robert H, and David S Stoffer. 2017. *Time Series Analysis and Its Applications: With R Examples*: Springer.

Song, Chaoming, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. "Limits of Predictability in Human Mobility." *Science* 327, no. 5968: 1018-1021. https://dx.doi.org/10.1126/science.1177170.

Song, Xuan, Hiroshi Kanasugi, and Ryosuke Shibasaki. 2016. *Deeptransport: Prediction and Simulation of Human Mobility and Transportation Mode at a Citywide Level*. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*.

Wallace, Gregory. 2020. "Airlines and Tsa Report 96% Drop in Air Travel as Pandemic Continues." Last modified 3:02 PM ET, Thu April 9, 2020. Accessed 07/30, 2020. Form of Item. https://www.cnn.com/2020/04/09/politics/airline-passengers-decline/index.html.

Weissman, Gary E, Andrew Crane-Droesch, Corey Chivers, ThaiBinh Luong, Asaf Hanish, Michael Z Levy, Jason Lubken, Michael Becker, Michael E Draugelis, and George L Anesi. 2020. "Locally Informed Simulation to Predict Hospital Capacity Needs During the Covid-19 Pandemic." *Annals of internal medicine*.

Wikipedia. 2020. "Floating Car Data." Last modified 8/17/2020. Accessed 10/3, 2020. Form of Item. https://en.wikipedia.org/wiki/Floating_car_data.

World Health Organization. 2020. "United States of America: Who Coronavirus Disease (Covid-19)." Accessed 07/30, 2020. Form of Item. https://covid19.who.int/region/amro/country/us.

Xu, Jun, Rouhollah Rahmatizadeh, Ladislau Bölöni, and Damla Turgut. 2017. "Real-Time Prediction of Taxi Demand Using Recurrent Neural Networks." *IEEE Transactions on Intelligent Transportation Systems* 19, no. 8: 2572-2581.

Xu, Lin, Yang Yue, and Qingquan Li. 2013. "Identifying Urban Traffic Congestion Pattern from Historical Floating Car Data." *Procedia-Social and Behavioral Sciences* 96: 2084-2095.

Yuan, Y. H., and M. Raubal. 2016. "Exploring Georeferenced Mobile Phone Datasets - a Survey and Reference Framework." *Geography Compass* 10, no. 6 (Jun): 239-252. https://dx.doi.org/10.1111/gec3.12269.

Yuan, Yihong, and Martin Raubal. 2012. *Extracting Dynamic Urban Mobility Patterns from Mobile Phone Data*. *International conference on geographic information science*: Springer.

Zhong, Chen, Stefan Müller Arisona, Xianfeng Huang, Michael Batty, and Gerhard Schmitt. 2014. "Detecting the Dynamics of Urban Structure through Spatial Network Analysis." *International Journal of Geographical Information Science* 28, no. 11: 2178-2199.

Zhong, Chen, Ed Manley, Stefan Mueller Arisona, Michael Batty, and Gerhard Schmitt. 2015. "Measuring Variability of Mobility Patterns from Multiday Smart-Card Data." *Journal of Computational Science* 9: 125-130.

Zhu, Lingxue, and Nikolay Laptev. 2017. *Deep and Confident Prediction for Time Series at Uber*. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*: IEEE.

Zhu, Xi, and Diansheng Guo. 2017. "Urban Event Detection with Big Data of Taxi Od Trips: A Time Series Decomposition Approach." *Transactions in GIS* 21, no. 3: 560-574.