Annals of the Association of American Geographers

ISSN: 0004-5608 (Print) 1467-8306 (Online) Journal homepage: http://www.tandfonline.com/loi/raag20

Modeling and Visualizing Regular Human Mobility Patterns with Uncertainty: An Example Using Twitter Data

Qunying Huang & David W. S. Wong

To cite this article: Qunying Huang & David W. S. Wong (2015) Modeling and Visualizing Regular Human Mobility Patterns with Uncertainty: An Example Using Twitter Data, Annals of the Association of American Geographers, 105:6, 1179-1197, DOI: 10.1080/00045608.2015.1081120

To link to this article: http://dx.doi.org/10.1080/00045608.2015.1081120

Published online: 23 Sep 2015.

Modeling and Visualizing Regular Human Mobility Patterns with Uncertainty: An Example Using Twitter Data

Qunying Huang* and David W. S. Wong[†,‡]

*Department of Geography, University of Wisconsin–Madison
[†]Department of Geography and Geoinformation Science, George Mason University
[‡]Department of Geography, University of Hong Kong

Traditional space–time paths show the spatiotemporal trajectories of individuals in one to several days. Based on data for such short periods, these space–time paths might not be able to show regular activity patterns, which are pertinent to various types of planning and policy analysis. Travel data gathered for longer periods might capture regular activity patterns, but footprints captured by these data also include irregular activities, introducing noises or uncertainty. Our objective is to determine the representative spatiotemporal trajectories of individuals, accounting for stochastic disturbances and spatiotemporal variability, but using activity data with longer duration. Therefore, we explore using Twitter data, which have relatively low and irregular spatial and temporal resolutions. This article introduces a methodology to construct individual representative space–time paths using various aggregation and spatiotemporal clustering techniques. To depict and visualize spatiotemporal trajectories with uncertain information, we propose space–time cones of variable sizes to reflect the spatial precision of the paths and use colors on the cones to represent the confidence level. To illustrate the proposed methodology, we use the geo-tagged tweets for an extended period. Our analysis indicates that the representative space–time path reasonably describes an individual's regular activity patterns. As visual elements, cones and cone colors effectively show the varying geographical precision along the path and changing certainty levels across different path segments, respectively. *Key Words: MTUP, regular activity patterns, representative space–time path, spatiotemporal variability and uncertainty.*

传统的时空路径, 展现出个人从一天到数日的时空轨迹。这些时空轨迹根据此般短期的数据, 或许无法展现出有关各种规划及政策分析类型的规律活动模式。较长时程中搜集的旅次数据, 或许能够捕捉规律的活动模式, 但这些数据所捕捉的足迹, 却仍包含不规律的活动, 引入了杂音或不确定性。我们的目标, 便是决定个人的再现时空轨迹, 并考量随机扰动和时空变异, 但是是运用较长时程的活动数据。因此, 我们运用相对而言具有较低且较不规律的时空分辨率的推特 (Twitter) 数据进行探讨。本文引介一个方法, 运用各种聚集和时空集合技术来建构个人的再现时空路径。为了描绘并可视化具有不确定信息的时空轨迹, 我们提出不同大小的时空锥形体, 用来反映这些路径的空间精确性, 并在锥形体上运用颜色来再现信赖区间。为了阐明提出的方法, 我们在延伸的时期中, 运用标注地理的推文。我们的分析指出, 再现的时空路径, 合理地描绘出一个人的规律活动模式。作为视觉的元素, 锥形体和锥形颜色分别有效地显示出路径上互异的地理精确性, 以及在不同路径部分中改变的确切性层级。 *关键词: 可调整时间单元问题, 规律活动模式, 再现的时空路径, 时空变异与不确定性。*

Las rutas tradicionales del espacio–tiempo muestran las trayectorias espaciotemporales de los individuos entre uno y varios días. Con base en datos para tan cortos períodos, estas rutas del espacio–tiempo podrían carecer de la capacidad de mostrar los patrones de las actividades regulares, las cuales son pertinentes a varios tipos de planificación y análisis de políticas. Los datos sobre viaje recogidos sobre períodos más largos podrían capturar, por otra parte, los patrones de las actividades regulares, aunque los rastros captados por estos datos también incluyen actividades irregulares, que agregan ruido o incertidumbre. Nuestro objetivo busca determinar las trayectorias espaciotemporales representativas de los individuos, tomando en cuenta perturbaciones estocásticas y variabilidad espaciotemporal, pero usando datos sobre actividades de una mayor duración. Consecuentemente, exploramos la utilización de datos de Twitter, que tienen resoluciones espaciales y temporales relativamente bajas e irregulares. Este artículo introduce una metodología para construir rutas individuales representativas del espacio–tiempo utilizando varias técnicas de agregación y agrupamiento espaciotemporal. Para representar y visualizar trayectorias espaciotemporales con información incierta, proponemos conos de espacio–tiempo de tamaño variable para reflejar la precisión espacial de las rutas, y usamos colores en los conos para representar el nivel de confianza. Para ilustrar la metodología que proponemos usamos los trinos geo-etiquetados para un período prolongado. Nuestro análisis indica que la ruta representativa del espacio–tiempo describe razonablemente los

Annals of the Association of American Geographers, 105(6) 2015, pp. 1179–1197 © 2015 by Association of American Geographers
Initial submission, November 2014; revised submissions, March and June 2015; final acceptance, June 2015
Published by Taylor & Francis, LLC.

patrones de actividad regular de un individuo. Por su condición de elementos visuales, los conos y los colores de los conos muestran de manera efectiva la variable precisión geográfica a lo largo de la ruta y los cambiantes niveles de certeza a lo largo de diferentes segmentos de la ruta, respectivamente. *Palabras clave: MTUP, patrones de actividad regular, ruta representativa del espacio–tiempo, variabilidad e incertidumbre espaciotemporal.*

The time–geographic framework introduced by Hägerstrand (1970) has been useful for analyzing human activity patterns with individual-based spatiotemporal data sets. Having a better understanding of activity patterns is not only important in studying transportation-related issues such as travel behavior (e.g., Kitamura 1988) and accessibility (e.g., Kwan 1998; National Research Council [NRC] 2002) but is now regarded as a critical aspect of inquiry in interdisciplinary studies of segregation and public health (e.g., Kwan 2013; Matthews and Yang 2013). Within the time–geographic framework, the space–time path (STP) has been used extensively to describe and analyze the activity patterns of individuals (e.g., Kwan 1999, 2000). A traditional STP is intended to show the spatiotemporal trajectory of an individual usually, although not necessarily, within a relatively restricted period, perhaps a day or several days (e.g., Kwan 2000, 2008; Chen et al. 2011). A path is composed of a series of tracking points connected by lines in a three-dimensional (3D) space (2D geographical space and 1D time on the vertical axis; Kwan 2004; Shaw, Yu, and Bombom 2008). These tracking points could indicate the start and end of an event spatiotemporally, and lines joining these points show that the individual either stays at the same location over a period or moves across locations within an event or among events. To a large degree, this STP depiction of movement adopts an event-based approach (Hornsby and Cole 2007; Long and Nelson 2013).

The traditional data source to construct STPs is often the travel diaries provided by recruited subjects willing to record or recall activities with detailed information over one to several days. With such data, STPs of individuals can be constructed and displayed in geographic information systems (GIS) to show the spatiotemporal trajectories for a specific day (or period) to support a variety of studies, including exploring differences in activity patterns among different population groups (e.g., Kwan 1999, 2000), determining whether individuals cross paths over a specific period (e.g., Buliung and Kanaroglou 2006; Yu 2007; H. Kang and Scott 2008), and evaluating interactions in a hybrid physical–virtual space (Ren and Kwan 2007; Shaw and Yu 2009). For many types of inquiry, however, such as urban and transportation planning, public health (e.g., MacPherson and Gushulak 2001; Bajardi et al. 2011), and tourism (e.g., Van der Knaap 1999; Lew and McKercher 2006), activity trajectories of a particular day might not be sufficient, as data collected for a day or two could be distorted by occasional or random activities and therefore might not capture long-term patterns or regular behaviors (González, Hidalgo, and Barabási 2008). Regular activity patterns are essential for planning, policy analysis, and decision making.

Although travel diaries provide relatively complete spatiotemporal recordings and information about the nature of trips, the advances in information and communication technology (ICT), particularly the pervasive adoption and penetration of location-aware technologies and devices in daily life, generate massive amounts of data. Some of these data are georeferenced (e.g., FourSquare and a portion of social media data provided by Twitter and Flickr). Different from travel diary data, these data are generated by users voluntarily and not intended to support activity pattern analysis. Although using these data for geographical and activity pattern research calls for careful consideration of data quality, they nonetheless capture some aspects of the spatiotemporal trajectories of the users. Moreover, they usually include a large number of users for relatively long periods as compared to the small numbers of travel diary survey subjects and the usual short time frames in conducting the surveys. In addition, the costs to acquire these social media data are minimal.

Despite their sparseness within a day, social media data cover longer periods and therefore could capture longer term patterns. The interday variability of spatial behavior should revolve around the longer term regular patterns. Therefore, an objective of this article is to explore whether geo-referenced data from Twitter can be used to construct STPs that depict the typical or representative activity patterns of individuals. Recognizing that such representative STPs might be derived from varying trajectories, a second objective is to incorporate such variability of activity patterns in depicting the representative STPs. This variability in patterns reflects the (un)certainty of the spatiotemporal trajectories.

In the next section, we highlight the limitations of existing STP and computational movement analyses in capturing regular activity patterns and the need to model movement data for deriving representative STPs. We then discuss the spatiotemporal uncertainty in determining the representative spatiotemporal trajectories. Next we describe our proposed approach and then illustrate our approach using Twitter data. The conclusion provides insights about the effectiveness of the approach and points out the potential of the approach in handling other movement data.

Space–Time Paths: What Type of Activity Patterns Do They Show?

Generalizability of Results of Current Space–Time Path Analyses

The current state of studies in STPs was partly attributable to the importance of understanding individual activity patterns in transportation planning (Kitamura 1988; Shaw and Wang 2000). Travel diary survey data, which provide comprehensive information about trips over a relatively short period, are often used to construct individual spatiotemporal trajectories visualized in 3D GIS environments. Several studies used two days of travel diary data to illustrate the effects of the perception of safety on one's travel behavior (e.g., Kwan 2008), to evaluate the variation in the STPs among different population subgroups (Kwan 1999, 2000), and to assess the activity patterns and intensity levels of a metropolitan area (Lee and Kwan 2011; Shen, Kwan, and Chai 2013). Studies have shown significant intrapersonal variability from day to day in travel patterns (Hanson and Huff 1981). Therefore, results obtained from these studies might not accurately reflect people's regular activity patterns over longer periods.

The STP of a particular day might not be sufficient to reflect an individual's typical activity pattern (Bayarma, Kitamura, and Susilo 2007), but "variability in the individual's travel pattern has a systematic, or nonrandom, component" (Hanson and Huff 1988, 111). Therefore, in deriving the "typical" travel pattern of an individual, both systematic variations and randomness should be included (Jones and Clark 1988; Pas 1988; NRC 2002). On the other hand, "an individual's daily activity and travel pattern varies from day to day, week to week, or season to season, in part because the individual and his or her household have needs that vary with time" (Bayarma, Kitamura, and Susilo 2007, 55). Thus, temporal variations in travel behavior for a specific individual could exist at different temporal scales (Janelle, Goodchild, and Klinkenberg 1988; Goodchild, Klinkenberg, and Janelle 1993), and typical activity patterns in daily life also change over time.

To derive typical activity patterns, activity trajectory data of individuals for very limited temporal coverage will not be useful as regularities in activity patterns of individuals emerge over an extended period. Regularities and variability in travel behavior can be explained by various individual-level socioeconomic and demographic conditions (Pas and Koppelman 1987; Hanson and Huff 1988; Jones and Clarke 1988; Pas 1988) and the nature of the trips (e.g., Hanson and Huff 1981; Koppelman and Pas 1985). These studies used travel records from several days to weeks to capture the regularities and variability, but many previous studies using STPs had relatively small sample sizes and short sample periods. An exception was a recent study by Wang and Sun (2015) that gathered travel data of individuals over 200 days, hoping that the data would reveal regular activity patterns. None of these studies explicitly employed STP to visualize regular activity patterns, however.

Space–Time Path Analyses

As Ellegård (1999) pointed out, after data recording activity patterns are collected, "the first question concerned how to similarly describe the varying everyday lives of different individuals, treating the great variation among the diaries in a proper way" (171). Our goal is to develop an approach to derive an STP for each individual to address this issue. Many analytical methods have been developed to analyze the similarity among individual paths (e.g., Kwan, Xiao, and Ding 2014). By comparing standardized STPs, which retain only the shapes of the original STPs but remove the absolute temporal and geographical references, common movement patterns can be revealed among multiple individuals (Kwan 1999). The standardized STP was among the set of tools (a total of six) implemented by Chen et al. (2011) for the Activity Pattern Analyst extension for ArcGIS to manage and analyze STPs. Comparing standardized paths could have limited significance, however, if these STPs are not representatives of people's regular activity and travel patterns and differences among these trajectories are largely random.

The relatively new field of computational movement analysis (CMA) focuses on analyzing detailed spatiotemporal trajectories (e.g., Gudmundsson, Laube, and Wolle 2012; Laube 2014; Yuan and Nara 2014), including the velocity of movements and their curvatures (e.g., Laube, Imfeld, and Wiebel 2005; Buchin, Buchin, and Gudmundson 2010). Many CMA studies evaluated the similarity among individual space–time trajectories (e.g., Sinha and Mark 2005), and some considered the geographic context in the analysis (e.g., G. Andrienko, Andrienko, and Heurich 2011; Buchin, Dodge, and Speckmann 2014). Aggregation is commonly used to visualize large volumes of trajectory data (e.g., Meratnia and de By 2002; G. Andrienko and Andrienko 2008; G. Andrienko et al. 2008; N. Andrienko and Andrienko 2011). Although aggregation is reasonably effective for summarizing a large number of links with similar origin–destination pairs, it is not effective for summarizing STPs, as the results cannot reflect the varying locations at different times of a day. Moreover, trajectories are aggregated as flows in 2D space, whereas SPTs are best shown in 3D space. Also, existing aggregation approaches do not consider statistical uncertainty (i.e., the extent to which the aggregated flows represent the original varying trajectories).

When Shaw, Yu, and Bombom (2008) clustered (aggregated) 3D STPs with similar trajectories to form the "generalized" STPs, the travel patterns of different population subgroups were aggregated, and paths for the same individuals were lost. Generalized paths also cannot reflect the (un)certainty due to the variability of trajectories. Demšar and Virrantaus (2010) proposed a volumetric kernel density approach to reflect the certainty of space–time trajectories of vessels in 3D space, but the data have to be dense enough to populate the relevant voxels; otherwise interpolation is required. Unfortunately, data from social media do not have the sufficient spatial and temporal densities required by the kernel density approach. In addition, results of this approach show intensity but not uncertainty.

Among the five challenges in movement analysis (Purves et al. 2014), two were to "(h)arness the full diversity of technologies to sense movement and its semantics" (2) and "(e)mbrace real but messy movement data" (3). Palmer et al. (2013) and other scholars advocated the use of mobile phone data for activity patterns studies (e.g., C. Kang et al. 2010; Järv et al. 2014; Silm and Ahas 2014), but none derived STPs of individuals to show their regular activity patterns. Shen, Kwan, and Chai (2013) categorized seven-day daily commuting patterns in Beijing, China, qualitatively but did not derive the typical trajectories.

Our objective of using social media data, particularly Twitter data, to derive the representative STPs that reflect typical activity patterns falls into the two challenges highlighted by Purves et al. (2014). Twitter data are not collected for our proposed research. These data are real and messy, but they do capture movement information.

Uncertainty in Representative Space–Time Paths

An estimated 1.5 percent of all tweets are geotagged (Murdock 2011). As with any social media data, using Twitter data has a few issues. Besides the selection bias that social media users are mostly young and technically savvy, Twitter data are relatively sparse spatiotemporally as compared to travel diary data. Data are gathered passively when users tweet, traveling or not. Therefore, using Twitter data in activity studies implicitly adopts "a coordinate-based representation of movement" (Long and Nelson 2013, 295). The path is a set of points (x, y) arranged in a temporal sequence with varied sampling and temporal granularities (Hornsby and Egenhofer 2002). This type of movement data can be characterized as "thin or shallow" as they capture only time and location, whereas the travel diary data are regarded as "thick or deep" as they include more detailed information (e.g., nature of events) that can be used to categorize trips (Hanson and Huff 1981; Koppelman and Pas 1985; Kwan 1999). Although Twitter data might not reflect the detailed trajectory of a user within a day, they offer selected locations of that individual over longer periods (several days to months; see Figure 1). Thus, by pooling the location information for multiple days, the longer temporal sample frames could compensate for the spatial sparsity of sample points in each day (Figure 1D).

Our general approach is to combine data from multiple days to derive a "generalized" trajectory or STP to depict the regular activity pattern of an individual over a certain period. We label our proposed STP as the representative STP to distinguish it from the generalized STP by aggregating paths of different individuals within a day (Shaw, Yu, and Bombom 2008).

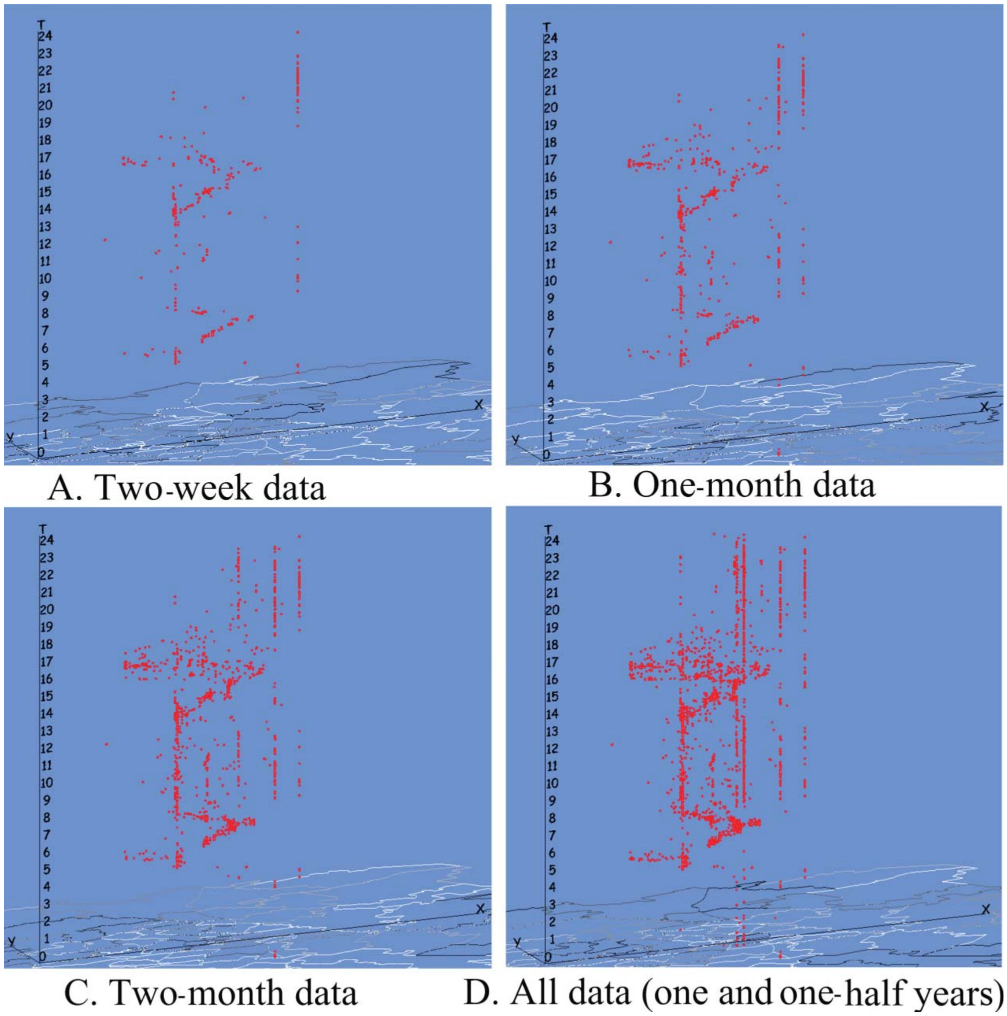


Figure 1. Trajectories of an individual over different periods. (Color figure available online.)

Space–time points (ST points) collected over a period can reflect regularities in activity patterns, but these regularities are partly composed of day-to-day variability (Hanson and Huff 1988). Part of the variability might also be the results of stochastic behavior creating noises (Bayarma, Kitamura, and Susilo 2007). These noises are deviations from regular activities or results of random behaviors. It is not possible, however, to determine whether an ST point is part of the regular pattern or just a random outcome with certainty. In other words, when a representative STP is derived using all ST points with highly varied patterns, this path is only accurate to a certain degree, as it includes some random points.

Therefore, besides depicting the regular activity pattern of an individual, another objective of this study is to model the uncertainty in the representative STP based on available data. Note that uncertainty used in this article is not the error in determining the location and time of an individual given some known locations as defined in movement analysis (e.g., Neutens et al. 2007; Long and Nelson 2013). Their focus was to estimate the likely locations of individuals from a modeling (e.g., Hornsby and Egenhofer 2002) or statistical perspective (e.g., Winter 2009; Winter and Yin 2010). Our interest in uncertainty focuses on the likelihood of the derived representative STP in depicting the individual's regular activity patterns given the interday variability and stochasticity in the data.

For most people, daily lives revolve around several major routes or activities, including work, school, rest, dining, entertainment, and social engagements (Hanson and Huff 1988; Pas 1988). For most of a day, these activities occur at certain "primary" locations (e.g., home, workplace), and occasionally they might be conducted at alternate locations. For instance, an individual might visit a particular restaurant or take a different route to work occasionally. Similarly, the individual might not conduct the same activities at the same times every day. These stochastic behavioral characteristics introduce spatial and temporal variability to the regular activity patterns.

Constructing Representative Space–Time Paths

Several assumptions are adopted in this study in deriving the representative STPs. We assume that each ST point records the time and location of that individual with reasonable accuracy. We adopt a coordinate-based approach (vs. an event-based approach), as our data do not capture information about events (Long and Nelson 2013). The data depict the regular activity patterns of individuals as ST points (i.e., a number of three-tuple of x, y, and t) over multiple days. Our objective is to determine the representative activity pattern of each individual in the twenty-four-hour period of a "typical" weekday by pooling ST points from multiple days. Although these ST points indicate the location of an individual, these spatiotemporal locations are the results of both regular activities and stochastic behavior. Thus, high frequencies of ST points in close proximity are likely associated with regular activities, whereas scattered ST points are likely associated with random activities. Therefore, our focus is those spatiotemporally clustered points. Practically, "aggregates" of these ST points rather than individual points are used to construct the representative paths.

Temporal Aggregation

Aggregation is a typical method for summarizing spatiotemporal data (e.g., G. Andrienko and Andrienko 2008). Assigning spatiotemporal points to different temporal windows or layers could be regarded as an aggregation process. This is analogous to controlling the temporal granularity. Hornsby and Egenhofer (2002) showed that using a coarser granularity generalizes the representation of data and provides a better depiction of the overall movement pattern. Using a fixed window size (e.g., one hour), time in the twenty-four-hour period is referred to as $t = 0, 1, 2, \ldots, 23$, and time slices are attached to the end of each temporal window. All points between 0:00 and 1:00 are assigned to the layer with time stamp t_0. Assuming that the temporal window size is uniform across layers, compressing ST points into temporal layers involves at least three parameters defining the temporal windows: size (or interval I), placement (T), and amount of overlap (L).

Using one hour as the temporal window size (I) is common, but finer or coarser granularities could be used. Given the high temporal variability relative to the typical travel diary data, using a window smaller than fifteen minutes might not be meaningful. On the other hand, a window larger than two hours will probably lump ST points of different activities. Given a particular window size (e.g., thirty minutes) starting times of windows could be placed differently. Starting at the hour or half-hour is typical, but the windows could be placed at the 15th or 45th minute of the hour; theoretically, it can be any time. But placing the windows at different times and using different sizes will produce different results, the problem known as the modifiable temporal unit problem (MTUP; Cheng and Adepeju 2014), the temporal analog of the modifiable areal unit problem (MAUP; Openshaw and Taylor 1981; Wong 2009).

To aggregate ST points over time, the moving time window method has been used (Shaw, Yu, and Bombom 2008) in which two consecutive windows overlap temporally. Analogous to the moving window operation in spatial analysis, temporal moving window operation smooths temporal variability, allowing the generalized temporal patterns to emerge. A parameter in this method is temporal overlap (L) between the two consecutive windows. Larger overlaps will produce smoother distributions, whereas using no overlap will not smooth the original data.

Therefore, window size or interval (I), temporal placement of windows (T), and the overlap (L) between two consecutive windows affect the temporal aggregation of ST points in deriving the representative STP of an individual. Using different parameter values will yield different results. One of the earlier methods to deal with the MAUP explored the sensitivity of results using different spatial resolutions and partitioning systems (Fotheringham and Wong 1991). Here, we use different temporal window sizes (I), placements (T), and sizes of overlap (L) to develop a more robust understanding of the derived activity patterns. Using smaller window sizes (I) will result in higher temporal density levels, and vice versa. Given each window size (I), windows can be placed at different times (T) over the hour, with temporal layers slicing the twenty-four hours at different times and with different sizes of overlap (L) between windows. Varying these parameter values will help evaluate the robustness of the resultant ST trajectories.

Handling Uncertainty Using Spatial Clustering

Compressing ST points into temporal layers reduces the temporal variability of data to uncover potential regularities. Clustering is used to address the spatial variability of the ST points to identify spatial trajectories (e.g., Shaw and Yu 2009; Chen et al. 2011). On each temporal layer, different points represent the varying locations of the individual within that temporal window. The varying locations are the results of both the regular activity patterns and random behavior deviating from the regularities. To describe the representative locations of an individual within a temporal window, cluster analysis could determine where points are concentrated, whereas locations associated with irregular activities could be discounted if they deviate greatly from the regular locations.

Within a temporal window, an individual might have more than one regularly visited location or multiple travel trajectories. Thus, multiple clusters could exist. Logically, the cluster capturing the largest proportion of activity points on the respective temporal layer probably best reflects the general location of the individual. This cluster, RC_t, could be regarded as the representative cluster for the temporal slice t (Figure 2). With a representative cluster on each temporal slice, the representative spatiotemporal activity trajectory for this individual can be constructed using these clusters over multiple layers.

If an individual varies greatly in his or her spatial behavior (systematic or stochastic variations), points on a temporal layer would be dispersed, and the resultant point cluster would have a relatively large spatial extent, implying that the data fail to pinpoint a more precise location for that individual during that temporal interval. If the individual had highly regular activity patterns, the representative clusters (RCs) would be compact. In other words, the individual was relatively confined to an area depicted by the highly clustered points within each temporal interval.

Although larger RCs indicate larger variations in locations, these larger clusters should not be perceived as unreliable if they were derived from a large number of activity points. Conversely, if an RC has a small spatial extent, it might not be reliable if it is formed only by a few points. Thus, RCs constructed by relatively more points should be more statistically significant than those constructed by fewer points. Assuming that n_t is the number of points in the temporal layer t and nR_t is the number of points forming the RC for temporal layer t, then the confidence level (CL) of that RC is defined as

$$CL_{R,\,t} = \frac{nR_t}{n_t}, \tag{1}$$

the proportion of points in layer t that forms the RC. When the CL of a particular RC is relatively low, it could imply the presence of one or more secondary clusters. There are other more sophisticated methods to measure the CLs of the RCs. The suggested method is nonetheless a reasonable starting point.

For most clustering methods, such as the popular K-mean algorithm in activity pattern analysis (Shaw and Yu 2009; Chen et al. 2011), the number of clusters has to be predetermined (Kanungo et al. 2002).

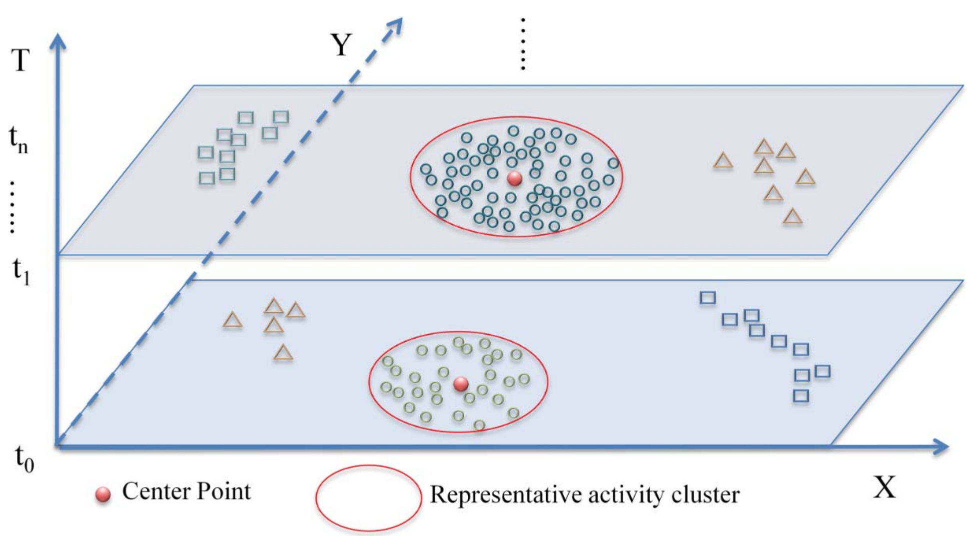


Figure 2. Modeling human mobility with a temporal layer model. (Color figure available online.)

Given that more than one cluster could exist within a temporal window, the number of clusters is best determined empirically by the data. Among various clustering algorithms, we chose DBSCAN, a popular data mining algorithm to detect unspecific shape clusters (Ester et al. 1996). DBSCAN does not specify the number of clusters in advance but uses the radius of a cluster (*Eps*) and the minimum number of points (*MinPts*) for a cluster as inputs.

Ester et al. (1996) showed that using a *MinPts* value smaller than four might misclassify random points as clusters, whereas a *MinPts* value ≥ 4 is unlikely to produce clusters of varying results. Therefore, we use a *MinPts* value of four. To determine the appropriate *Eps* value, we adopted the interactive procedure suggested by Ester et al. (1996). DBSCAN was first executed using a predefined arbitrary distance, adjusted subsequently by the process, and *MinPts* was set to four, likely producing multiple clusters. Clusters with less than the predefined relatively small number of points (e.g., fifty) are considered as noise and removed. For every point p_{ij} in each remaining cluster (C_i), the distance from p_{ij} to its *MinPts*th nearest neighboring point Dis_{ij} is derived. The average *MinPts*th nearest neighbor distance for all points in cluster C_i is calculated as follows:

$$AMinP_{tsi} = \frac{\sum_{j=1}^{n_i} Dis_{ij}}{n_i}, \tag{2}$$

where n_i is the number of points in cluster i. Finally, the maximum $AMinPts_i$ among all clusters is chosen as the final *Eps* value.

Handling Uncertainty Using Temporal Clustering

Although the spatial clustering process helps remove dispersed locations resulting from random activities, the temporal distances between ST points were not considered in the process. As the spatial clustering process compresses all ST points within each temporal window onto a time slice, an ST point could be close to other points within an RC spatially but is an outlier temporally. Therefore, points within the RC should be reexamined to ensure that they are temporally representative of the activity patterns. The first step is to "restore" the original time stamps of ST points constituting an RC. The ST points that are relatively far from other points temporally could be the results of random activities. A temporal clustering process will be used to remove these temporal outliers.

An RCs temporal window can be divided into finer temporal frames and ST points of the RC spread across these temporal frames. The frame with the largest number of points, which implicitly captures the regular activity patterns, is merged with adjacent temporal frames incrementally to form a temporal cluster (TC) until the proportion of points in the TC to the total number of points in the RC reaches a predefined threshold, k. For example, the seventh frame with the most points (11) is the initial frame of a TC (Figure 3). Adjacent frames are incrementally added to the TC until the proportion of points included in the TC reaches 70 percent (assumed k) of that in the RC. These frames constituting a TC form a temporal segment. This algorithm is inspired by the classic Dijkstra's algorithm to determine the shortest path between vertices in a graph with nonnegative edges.

This method involves the RC in each layer and the three following parameters: window size or interval (I), size of frames (f) within each interval, and k. These parameters can be adjusted during visualization. Normally, f is much smaller than I. If I is relatively large, each interval might include loci of multiple regular activities such that the RCs might not provide reasonably precise geographical footprints of the individuals. On the other hand, if I is too small, ST points for the same regular activity could be sliced into different intervals and thus the associated ST points might not form a cluster. Size of frames is relatively unimportant:

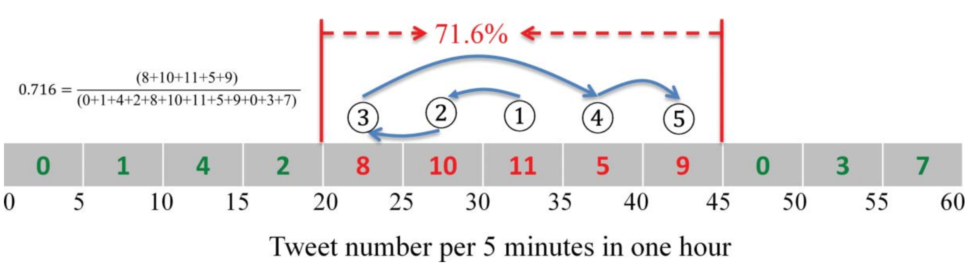


Figure 3. Time segment detection. (Color figure available online.)

Too large a frame size could lead to coarser TCs, and too small of a frame size will require more steps to determine the clusters. Eventually, trial and error might determine their sizes. As threshold k is the proportion of ST points in an RC to be retained to determine its associated TC, a smaller k removes more temporally dispersed ST points.

Given a temporal layer t, the TC includes only a portion of points in the RC that meets the threshold k. Therefore, the CL of the TC for temporal layer t, which is dependent on k and $CL_{R,t}$, is

$$CL_{T,\,t} = k * \ CL_{R,\,t}. \tag{3}$$

The ST points represent the locations of both regular and stochastic activities. Our premise is that large numbers of ST points in close spatiotemporal proximity reflect regular activities, and only these clustered points should be used to depict the representative STP. The $CL_{R,t}$ of an RC and k of a TC reflect the levels of strictness in removing the locations of random activities. Lower $CL_{R,t}$ values imply that only those points in close spatial proximity are retained in the RCs, but the more spatially dispersed points, likely representing irregular activities, are removed. These lower values also mean that the resultant RCs might not capture all locations of regular activities. With higher $CL_{R,t}$ values, the RCs are more likely to include locations of regular activities. Similarly, by lowering k for TC, temporally dispersed locations of random behavior will likely be moved, raising the risk of not capturing all locations of regular activities. Although these CLs do not necessarily correlate with the spatial extents of the clusters, higher $CL_{R,t}$ will likely produce larger primary clusters but less precise representative STPs. If $CL_{R,t}$ is raised further in determining RCs, secondary clusters could be included.

Visualizing Representative Space–Time Trajectories

Many studies have discussed using STPs to display spatiotemporal patterns of individuals' daily activities (e.g., Kwan 2000, 2004; Buliung and Kanaroglou 2006; Ren and Kwan 2007; Shaw and Yu 2009; Chen et al. 2011; Shen, Kwan, and Chai 2013). Using only line segments to connect different ST points of events is not sufficient to meet our objective of capturing the spatiotemporal variability of the representative STPs, however. In addition, Twitter data cannot support such event-based visualization as the data capture only the spatiotemporal coordinates of individuals without event information (e.g., we cannot specify when a particular event begins or ends).

After ST points are segmented into temporal layers and compressed temporally to temporal slices, RCs are derived for temporal windows with specific intervals (I), placements (T), and amounts of temporal overlap (L). The RCs across different temporal layers should include locations representing places or tracks associated with the individual's regular activities, likely revolving around home, workplace, or school. Connecting these RCs in a temporal sequence produces a form of an individual's activity path. Each RC consists of a set of points from which a circumscribed circle can be derived to represent the cluster geometrically and spatially. Several attributes of the circle and the corresponding RC should be derived: circle centroid, circle radius that indicates the spatial extent of the RC, CL of the RC (Equation 1), and time stamp of the temporal layer.

Different RCs have circles of different sizes or radii, representing the uncertainties of locations in the respective periods. To represent the activity path over different periods, circles of different sizes representing the RCs can be connected by variable-sized cylinders or truncated cones. A truncated cone uses the circumscribed circles of two consecutive RCs as the top and bottom, and the conic surface connects the top and bottom circles (Figure 4). The centers of the two circles can be in different locations producing a slanted cone. The conic segment is labeled as a space–time (ST) cone in this specific spatiotemporal analytical context. Conic shapes were also used in trajectory studies by Hornsby and Egenhofer (2002), Winter (2009), and Winter and Yin (2010), but they were constructed using different principles.

Across all temporal layers, RCs are connected by ST cones to depict the representative STP (Figure 4). Larger cones reflect that the individual was in various locations regularly in the respective periods, and smaller cones reflect that the individual was confined to smaller areas. The CL (Equation 1), however, which indicates the likelihood that the individual was within the cones of respective temporal layers, is indicated by color hues on the conic surface as a visual variable. As each CL refers to the RC derived from ST points over a temporal window (e.g., t_{i-1}, which covers the period from the $i-1$th hour to the ith hour), the CL is applicable to the time between the two temporal slices of t_{i-1} and t_i (Figure 4), and the conic surface will be colored

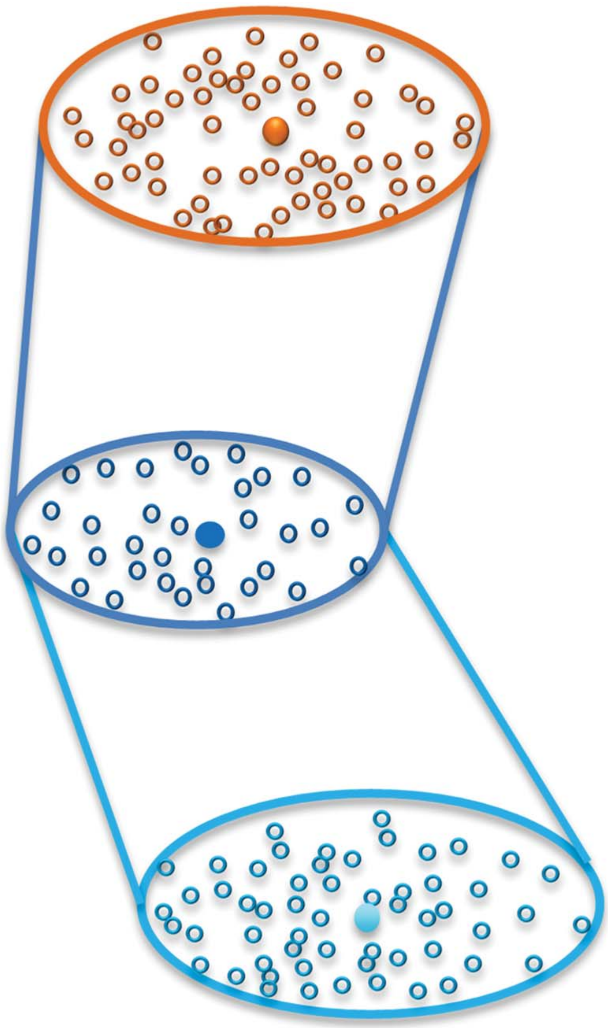


Figure 4. Constructing variable size space–time cones between two time periods. (Color figure available online.)

according to $CL_{R,t}$ when $t = i - 1$. Again, larger sized RCs do not necessarily have higher CLs, and a relatively small RC can have a relatively high CL if the individual has a highly consistent activity pattern.

The representative STPs depicted by ST cones are derived from spatial clusters that account for spatial variability. ST points in each RC are further clustered temporally with threshold proportion k to derive a TC. Each TC is not attached strictly to a time slice but is stretched across several frames within each temporal interval. Each TC has fewer points than the corresponding RC (assuming that k < 100 percent). Using these points, we can derive a circumscribed circle to represent the spatial extent of the TC and use the centroid of this circle to represent the general location of the TC. The circle should not be attached to a temporal slice or frame (t_0, t_1, t_2, . . ., t_n), however, as a TC is composed of points over several frames around the initial frame. Therefore, we place the circles representing the TCs at the starting times of the initial frames. Using these circles, we derive the ST cones (Figure 5), interpreted in the same manner as those cones for RCs, but the clusters are more precise temporally within each window.

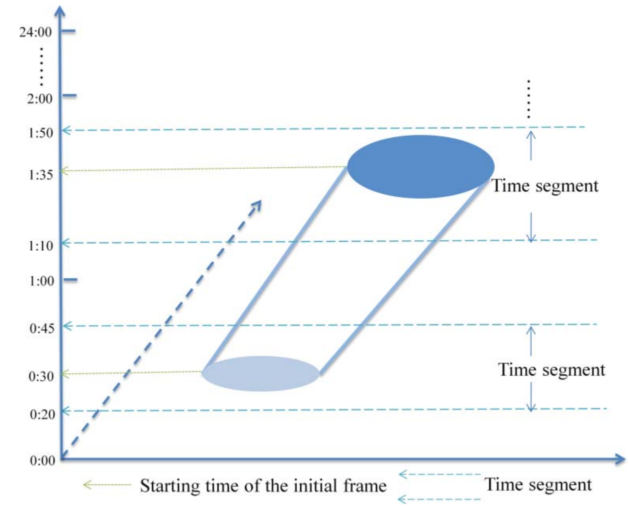


Figure 5. Space–time cone paths accounting for temporal uncertainty. (Color figure available online.)

The CL of each TC, $CL_{T,t}$ (Equation 2), is visualized by color hues on the conic surface, but as $CL_{T,t}$ combines both the CL of the corresponding RC and the threshold proportion k, the CLs of TCs show the likelihood that the individual's systematic activity tracks are found inside the prescribed areas during certain frames within each temporal window.

To handle the MTUP effects in determining the RCs and TCs, representative STPs were derived using a range of parameter values for window size (I), window placement of initial time (T), frame size (f), and size of overlapping windows (L) between two adjacent periods. Our goal is to reveal robust spatiotemporal patterns associated with the regular activities. Spatiotemporal tracks from random activities can be removed when various parameter values are used. These parameters, in combination with the CL to determine RC and the threshold level k to determine TC, control the processing and visualization of the multiple-day ST points depicting the representative STP of an individual.

An Example Using Twitter Data

To illustrate how the proposed methodology can be used to derive and visualize representative ST

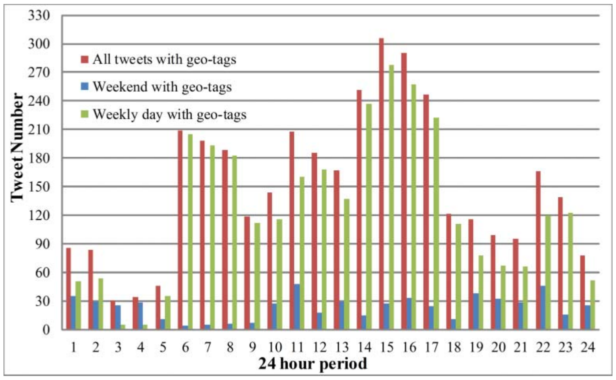


Figure 6. Geotagged tweet statistics for each hour period for the selected user. (Color figure available online.)

trajectories of individuals over multiple days, we use tweets with geotags. Using Twitter's streaming application programming interface (API), we harvested geotagged tweets within the United States for an extended period. To select appropriate subjects for our study, we first randomly selected several Twitter users with relatively large numbers of geotagged tweets and manually examined their online profiles to make sure that they were not organizations. As a result, a user in Austin, Texas, was chosen to demonstrate our methods of modeling and visualizing individual STP. We also retrieved three followers (friends) of this user in the same city for the purpose of multiple-user visualization. Therefore, these selected individuals did not represent any specific segment or group of the population. The primary user in our analysis posted 3,611 geotagged tweets between 11 November 2013 and 13 April 2014. Figure 6 shows the hourly distribution of tweets throughout this period. Most tweets (46 percent) were posted between 10:00 and 17:00. This user was especially active between 14:00 and 16:00 (23 percent of total tweets); messages were rarely posted between 0:00 and 4:00. Also, this user tweeted much less frequently during the weekends (Figure 6). A tool has been developed with C++, using the Geospatial Data Abstraction Library (GDAL) to handle geospatial data and OpenGL for 3D rendering.

Representative ST Paths

The twenty-four-hour spatiotemporal footprints of this individual were compressed across weekdays over the study period in three different formats in a 3D perspective view (Figure 7), not a horizontal cross-sectional view. The vertical dimension (T) of the three-graph panel represents the twenty-four hours. In Figure 7A, locations extracted from the geotagged tweets were mapped to the geographic coordinates but were elevated to the times (vertical dimension) when the tweets were sent. Spatiotemporal points are displayed in the same color within each one-hour period. Although certain degrees of clustering of these ST points are visible, obvious spatiotemporal patterns are not recognized, except that this individual was clearly in different locations (from 5:00–17:00) and was relatively confined to specific locations in early mornings

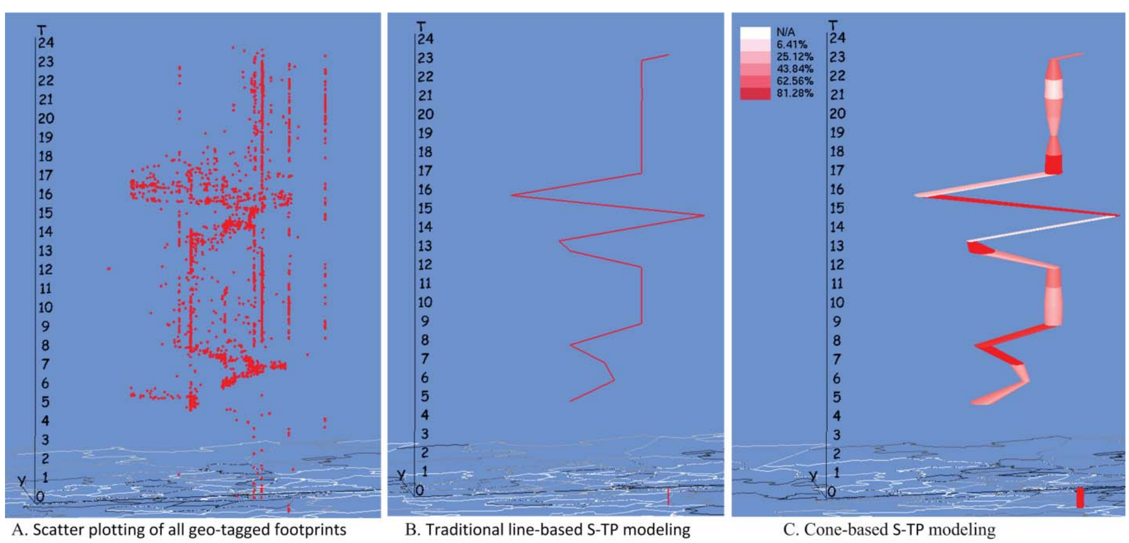


Figure 7. A comparison of visualization results of traditional space–time (ST) line paths and ST cone paths with one hour as the time interval. (Color figure available online.)

and evenings. These ST points have relatively high degrees of spatiotemporal variability, reflecting both the variation in regular activity patterns and random movements.

Using four points as the minimum and the distance determined by the procedure described earlier about clustering, we identified the RC for each temporal window. Centroids of the circumscribed circles of RCs were connected by line segments to produce the STP, indicating the individual's general locations throughout a typical weekday (Figure 7B). Similar to the traditional STP depiction, this path cannot reflect the variability of activity patterns. The representative STP based on the loci of RCs is shown using ST cones (Figure 7C). These ST cones are underpinned by a wealth of information. The variable cone sizes depict

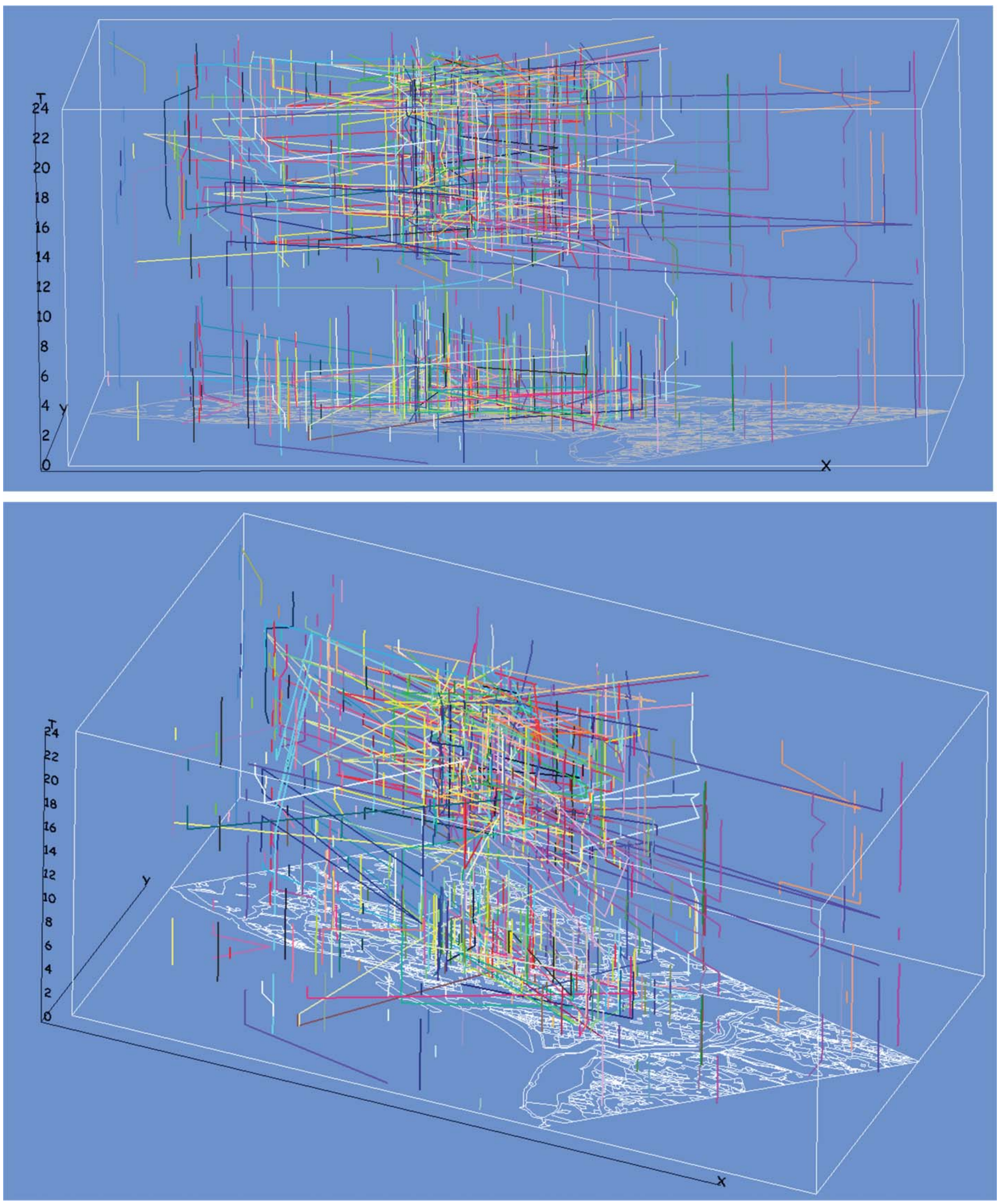


Figure 8. Representative space–time paths (STPs) of 1,986 Twitter users in Washington, DC, depicted by the traditional space–time (ST) line segments in different 3D perspective views. (Color figure available online.)

the spatial variability of activity patterns in different periods. For example, this individual had relatively moderate to high variability in activity patterns from 9:00 to 12:00, 17:00 to 18:00, and 21:00 to 22:00, as indicated by the relatively wide cones; but this individual had highly regular activity patterns between 14:00 and 16:00 involving at least two relatively distance locations. The narrow cones between 14:00 and 16:00 indicate that this individual was confined in those restricted locations. Given a dearth of sample points between 1:00 and 5:00, no representative clusters were identified, so no ST path was produced, but an ST cone with the size of RC for 0:00 is used to depict the trajectory between 0:00 and 1:00.

The different cone colors, the CLs of the respective representative clusters, show the likelihoods that the individual was in the locations depicted by the cones. The CLs range from 6.41 percent to 81.28 percent (Figure 7C). According to cone colors, the individual was likely in the locations depicted by the cones between 17:00 and 18:00 but less likely in those locations between 9:00 and 11:00 and 21:00 and 22:00.

Between 5:00 and 6:00, this individual had activities in a variety of locations regularly so the cone for the STP was larger but also captured relatively few ST points (Figure 7C). Accordingly, the CL of this individual's general locations during this period is lower. But between 6:00 and 7:00, the activity locations were more consistent and thus the CL was raised. Although we cannot describe and explain all ST segments, the general idea is that CLs are based on the proportions of ST points captured by the cones (or RCs) in different temporal intervals, whereas the sizes of cones are determined by the spatial variability of activity locations.

Although the proposed representative STPs have the same structural characteristics as those of the traditional STPs derived from the travel diary data, the two types of STPs serve different purposes. As the traditional STPs are based on detailed but short-term data,

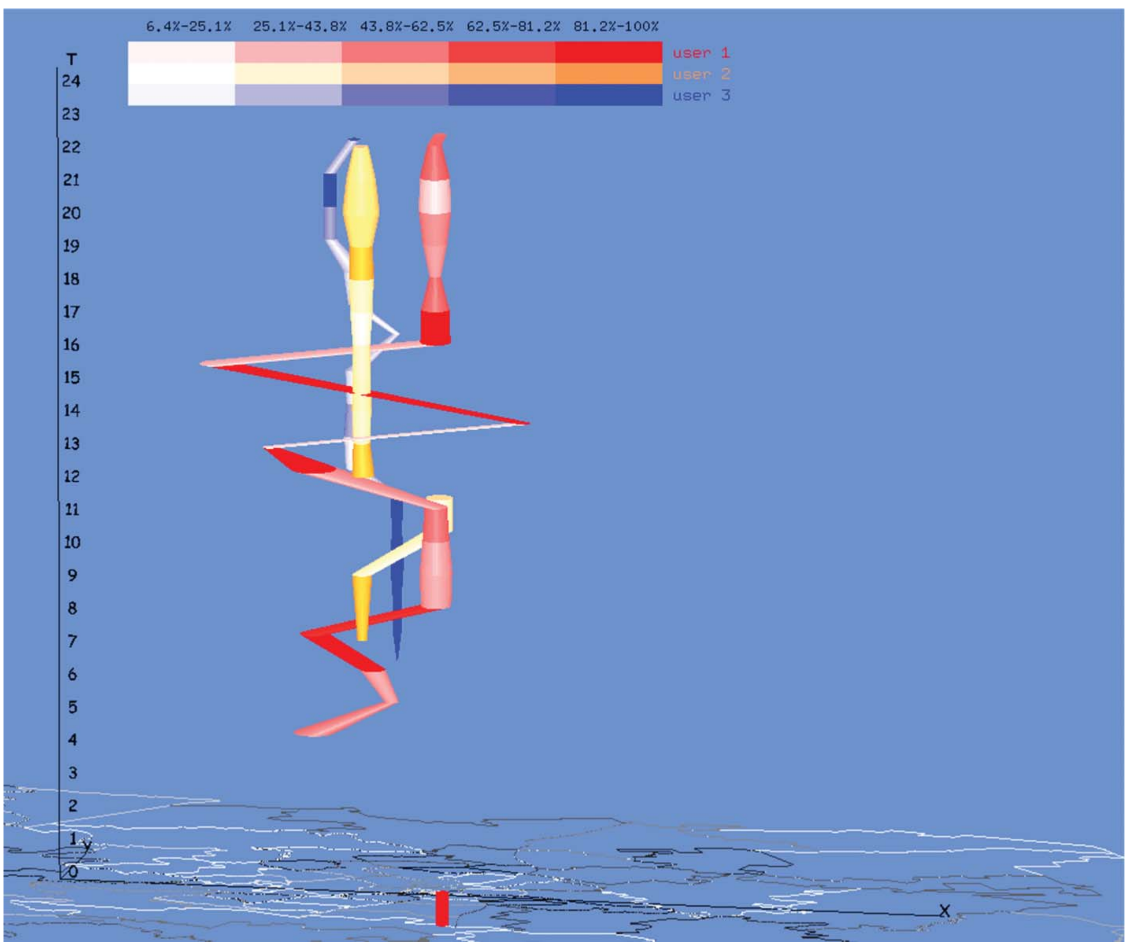


Figure 9. Representative paths in space–time (ST) cone for three Twitter users in Austin, Texas. Three colors are for the three users, and different hues indicate various confidence levels. (Color figure available online.)

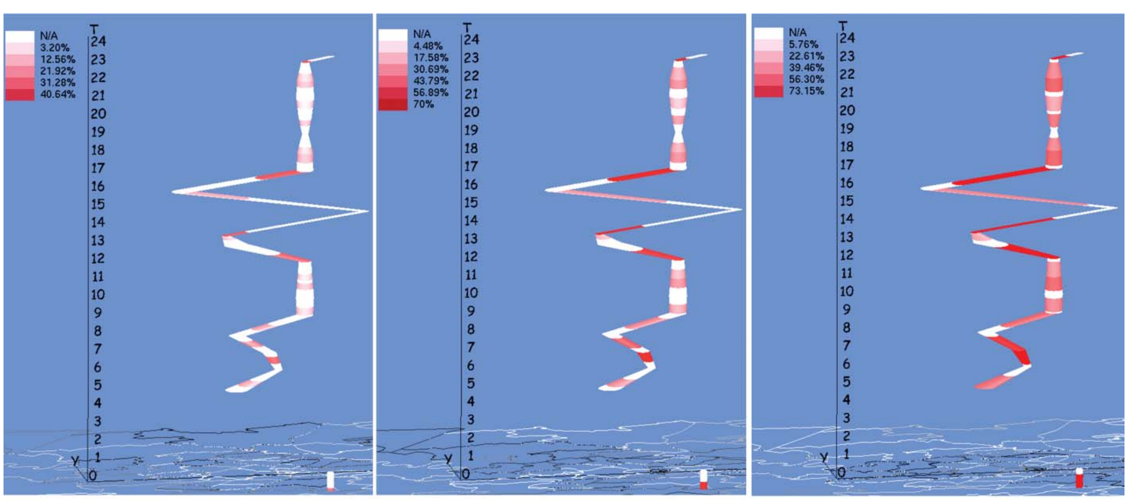
they might not be reliable for planning, policy formulation, and decision making. The representative STPs are intended to show regular activity patterns and are critical for planning and decision making. For demonstration purposes, the representative STPs of 1,986 Twitter users who claimed to reside in Washington, DC, in their profiles[1] are illustrated (Figure 8). These users were active in two periods—the morning and afternoon—and the spatiotemporal commuting patterns were revealed. The trajectories of this large number of users were based on an extended period showing regular activity patterns, which are useful for planning and public management.

These representative STPs can be analyzed by multiple methods, including the standardized STPs intended for the traditional STPs (Kwan 1999; Chen et al. 2011). Such analyses would not be correct in studying the spatiotemporal interaction among individuals, however, as the variability of STPs was not considered. As an example, the representative STPs of three users in the Austin, Texas, area show high spatial variability during the day, conforming to the general expectation of higher mobility levels in daily hours (Figure 9). To correctly analyze the spatiotemporal interaction among individuals, representative ST cones should be used to account for the spatial variability.

ST Cones with Temporal Uncertainty

To account for the temporal variability of ST points, only ST points in close temporal proximity within RCs should be used to construct the TCs. The results at three different threshold levels (k) of 50 percent, 70 percent, and 90 percent using a one-hour window size are provided (Figure 10). The overall shapes of the ST cones are very similar to the cones based on only RCs (Figure 7C), whereas the CLs of TCs (or segments in Figure 10) take into account the temporal variability and are always less than the CLs of their corresponding RCs, which do not consider temporal variability. In addition, because only a portion of ST points (inside a window) are used to form a TC, the TC's CL is only applicable to the corresponding portion of that temporal interval (i.e., sections of cones outside the TC are colored in white). As a TC is constituted by fewer points than the corresponding RC, the cone sizes of TCs are also smaller than those of RCs. The circumscribed circles for TCs are placed at the beginning of the initial frames rather than at the starting time. Therefore, the ST cones in Figure 9 and Figure 7C are different at a fine scale but overall look similar. Different from CMA, which estimates the likely locations given some movement trajectories, the proposed methodology cannot predict locations among ST points due to the sparseness of data and the fact that data from multiple days are pooled. To be conservative, time segments with no data are colored in white, indicating insufficient data to construct the respective ST segments.

Using a threshold (k) of 70 percent (Figure 10B), there is a 70 percent confidence level that the individual was within the temporal frames of TC with a mean time segment of twenty-five minutes. Using higher thresholds results in longer time segments (more temporal frames), indicating higher certainty but less



A. k = 50% B. k = 70% C. k = 90%

Figure 10. Trajectories with one hour as time window size but different time segment thresholds. (Color figure available online.)

precision temporally about the individual's locations within each time interval. For example, when the threshold is 90 percent, the average time segment is thirty-seven minutes, and many segments displayed in red (or dark gray in the grayscale figure) are spread across almost the entire one-hour temporal window (Figure 10C). Although only three threshold levels were evaluated, one could select an average time interval as a threshold to control the temporal uncertainty during the visualization.

Modifiable Temporal Intervals and Placements

To address the MTUP, varied window sizes were used in determining the RCs and TCs to produce different portrayals of the representative STPs. Three STPs using three window sizes were investigated: one-half hour (Figure 11A), one hour (Figure 11B), and two hours (Figure 11C). Although the three paths display similar overall trajectories, the path using the smallest window shows the greatest details, but the path using the two-hour window is the smoothest with the fewest details. Smoother paths using larger windows also produce larger cones, meaning that the locations of the individual cannot be known precisely in space. With larger windows, the individual was likely to move among more locations, creating representative clusters with larger spatial extents. Thus, the confidence levels for paths derived from larger windows are lower (note that the same colors for the three paths in Figures 11A, 11B, and 11C refer to different probabilities). Using a larger window size (two hours), the RCs between 1:00 and

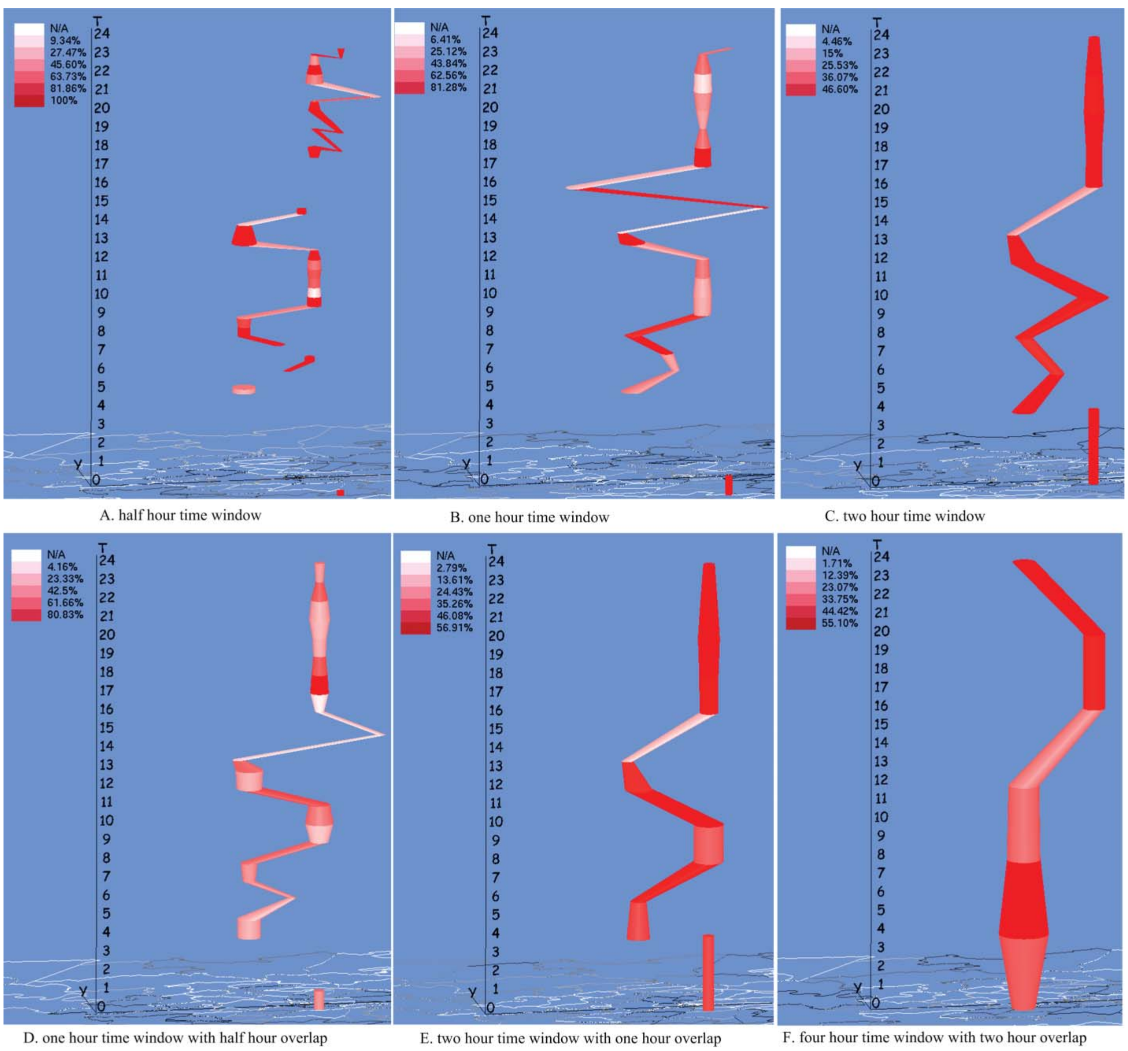


Figure 11. Cone-based space–time paths of the selected individual using different temporal window sizes and different degrees of window overlaps. (Color figure available online.)

4:00 are detected even when the ST points in that period were relatively sparse.

Using overlapping windows is another method to handle the spatiotemporal variability of representative STPs. The resultant STPs, corresponding to the window sizes in Figures 11A, 11B, and 11C, in series, but with one-half-hour (11D), one-hour (11E), and two-hour (11F) overlaps are illustrated in the corresponding Figure 11 panels. Compared to the STPs without overlaps (Figures 11A, 11B, and 11C), the use of overlapping windows produces smoother paths and results in more representative clusters. In addition, cone sizes with overlapping windows are wider along the trajectory paths, indicating that the loci of the STPs with overlapping windows are less precise spatially than their counterparts without overlapping windows.

Different placements of the windows also influence how ST clusters representing activities are sliced into different temporal layers, therefore resulting in highly varied STPs. Figure 12 shows the results using different placements (fifteenth, thirtieth, forty-fifth, and sixtieth minute) and two window sizes (one and two hours). With the one-hour window, the first temporal layer (T_0) includes all points between 23:15 and 0:15 with the fifteenth minute as the placement (Figure 12A) and between 23:30 and 0:30 with the thirtieth minute as the placement (Figure 12B), and so on.

The general notion that using different placements produces different space–time trajectory patterns is confirmed (Figure 12). With different window sizes, not only are the spatial precisions of the trajectories different (i.e., larger windows with lower precision) but the segments' confidence levels are different. Given the same window size, however, the overall space–time trajectory with different placements is evident. Whereas this multiscale (window size) and variable placement approach generates varying results, if consistent patterns across results can be identified, these patterns should be very reliable. For instance, most results show highly similar trajectories of this individual between 10:00 and 14:00, demonstrating that the trajectories are reliable even under the MTUP.

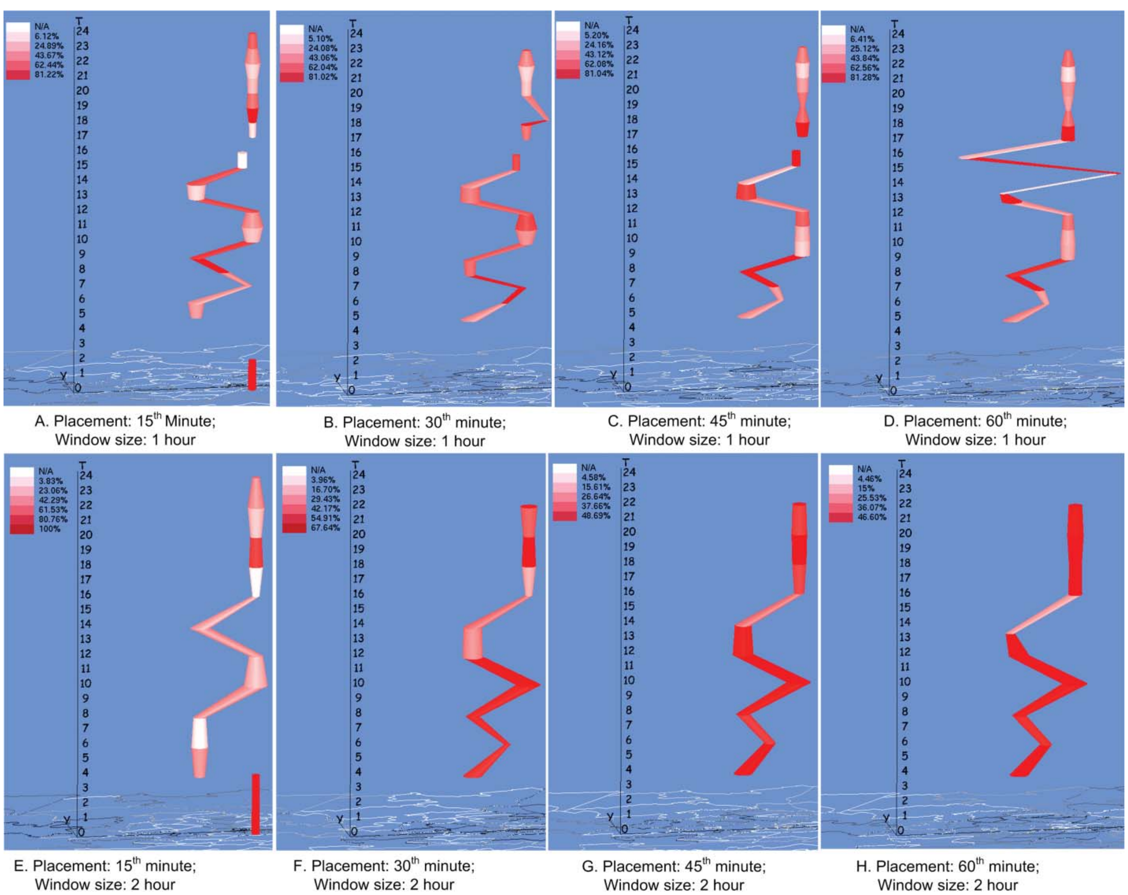


Figure 12. A comparison of visualization results of different placements (fifteenth, thirtieth, forty-fifth, and sixtieth minute) and time window sizes (one and two hours). (Color figure available online.)

Conclusion and Discussion

We introduce a methodology for deriving representative STPs from multiple-day movement data for different individuals. The proposed methodology addresses the spatiotemporal variability of an individual's movement with irregular spatiotemporal resolutions to capture the varying regular activity patterns and isolate the stochasticity of random travel behavior. The proposed methodology includes a visualization protocol that depicts the representative STPs with uncertainty information. In contrast to the traditional STPs, which describe the mobility patterns of individuals in one or a few survey days, the representative STP methodology depicts regular or systematic activity patterns, which are more relevant to many planning decisions and policy formulations.

By processing the ST point data over a range of parameter settings, varying depictions of representative STPs were produced, but consistent patterns across these varying depictions reveal the regular activity patterns. Although the focus is identifying spatial and temporal clusters meeting certain confidence levels or probability thresholds, the proposed methodology allows for interactive manipulation of the rendering of representative STPs by modifying various parameters, including the CL and threshold. Therefore, various CLs can be explored to evaluate changes in path depiction. By lowering CLs, secondary spatial clusters could be included in the representative STPs. The proposed modeling and visualization methodology is intended to handle ST points of individuals for multiple days but can address movement data other than humans (e.g., vehicles, animals). Moreover, the methodology is applicable to recurring movement data that capture the entire trajectories using the Global Positioning System or recalls or recording in travel diaries.

An objective in STP studies is to analyze the activity patterns of individuals. GIS tools have been developed to support such endeavors based on 3D representation of lines and nodes of the STPs (e.g., Yu 2007). Although one might question the usefulness of studying the activity patterns of specific days, the analytical tools developed to analyze traditional STPs will not be entirely applicable to representative STPs introduced here. Although whether two individuals "share" a space regularly is probably more meaningful than if individuals run into each other accidentally (e.g., from a planning perspective), analyzing regular activity patterns depicted by representative STPs will require an enhanced set of tools that can handle the spatiotemporal variability of activity patterns and the probabilistic nature of the representative STPs.

Note

1. The current prototype takes about six seconds to derive and render the representative ST path and associated ST cones for one user. Deriving and rendering of these paths and cones for a large number of users is therefore time-consuming, as the computing cost of the DBSCAN algorithm is O ($n4$) where n is the number of clustered points. Advanced computing methods including parallel computing, cloud computing, and graphics processor unit computing will be explored to support the proposed framework.

References

Andrienko, G., and N. Andrienko. 2008. Spatio-temporal aggregation for visual analysis of movements. In *IEEE Symposium on Visual Analytics Science and Technology, 2008 (VAST'08)*, 51–58. New York: IEEE.

Andrienko, G., N. Andrienko, and M. Heurich. 2011. An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science* 25 (9): 1347–70.

Andrienko, G., N. Andrienko, I. Kopanakis, A. Lightenberg, and S. Wrobel. 2008. Visual analytics methods for movement data. In *Mobility, data mining and privacy: Geographic knowledge discovery*, ed. F. Giannotti and D. Pedreschi, 375–409. Berlin: Springer-Verlag.

Andrienko, N., and G. Andrienko. 2011. Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics* 17 (2): 205–19.

Bajardi, P., C. Poletto, J. J. Ramasco, M. Tizzoni, V. Colizza, and A. Vespignani. 2011. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS ONE* 6 (1): e16591.

Bayarma, A., R. Kitamura, and Y. O. Susilo. 2007. Recurrence of daily travel patterns: Stochastic process approach to multiday travel behavior. *Transportation Research Record* 2021:55–63.

Buchin, K., M. Buchin, and J. Gudmundson. 2010. Constrained free space diagrams: A tool for trajectory analysis. *International Journal of Geographical Information Science* 24 (7): 1101–25.

Buchin, M., S. Dodge, and B. Speckmann. 2014. Similarity of trajectories taking into account geographic context. *Journal of Spatial Information Science* 9:101–24.

Buliung, R. N., and P. S. Kanaroglou. 2006. A GIS toolkit for exploring geographies of household activity/travel behavior. *Journal of Transport Geography* 14 (1): 35–51.

Chen, J., S. L. Shaw, H. Yu, F. Lu, Y. Chai, and Q. Jia. 2011. Exploratory data analysis of activity diary data: A space–time GIS approach. *Journal of Transport Geography* 19 (3): 394–404.

Cheng, T., and M. Adepeju. 2014. Modifiable temporal unit problem (MTUP) and its effect on space–time cluster detection. *PLoS ONE* 9 (6): e100465.

Demšar, U., and K. Virrantaus. 2010. Space–time density of trajectories: Exploring spatial-temporal patterns in movement data. *International Journal of Geographical Information Science* 24 (10): 1527–42.

Ellegård, K. 1999. A time-geographical approach to the study of everyday life of individuals: A challenge of complexity. *GeoJournal* 48 (3): 167–75.

Ester, M., H. P. Kriegel, J. Sander, and X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD* 96, 226–31. Palo Alto, CA: AAAI Press.

Fotheringham, A. S., and D. W. S. Wong. 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23:1025–44.

González, M. C., C. A. Hidalgo, and A. L. Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453 (196): 779–82.

Goodchild, M. F., B. Klinkenberg, and D. G. Janelle. 1993. A factorial model of aggregate spatio-temporal behavior: Applications to the diurnal cycle. *Geographical Analysis* 25 (4): 277–94.

Gudmundsson, J., P. Laube, and T. Wolle. 2012. Computational movement analysis. In *Springer handbook of geographic information*, ed. W. Kresse and D. M. Danko, 725–39. Berlin: Springer-Verlag.

Hägerstrand, T. 1970. What about people in regional science? *Papers of the Regional Science Association* 24 (1): 7–21.

Hanson, S., and J. O. Huff. 1981. Assessing day-to-day variability in complex travel patterns. *Transportation Research Record* 891:18–23.

———. 1988. Systematic variability in repetitious travel. *Transportation* 15:111–35.

Hornsby, K. S., and S. Cole. 2007. Modeling moving geospatial objects from an event-based perspective. *Transactions in GIS* 11 (4): 555–73.

Hornsby, K. S., and M. J. Egenhofer. 2002. Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence* 36:177–94.

Janelle, D. G., M. F. Goodchild, and B. Klinkenberg. 1988. Space–time diaries and travel characteristics for different levels of respondent aggregation. *Environment and Planning A* 20:891–906.

Järv, O., K. Müürisepp, R. Ahas, B. Derudder, and F. Witlox. 2014. Ethnic differences in activity spaces as a characteristic of segregation: A study based on mobile phone usage in Tallinn, Estonia. *Urban Studies*. Advance online publication. doi: 10.1177/0042098014550459.

Jones, P., and M. Clarke. 1988. Significance and measurement of variability in travel. *Transportation* 15:65–87.

Kang, C., S. Gao, X. Lin, Y. Xiao, Y. Yuan, Y. Liu, and X. Ma. 2010. Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In *Proceedings of the 18th International Conference on Geoinformatics*. New York: IEEE.

Kang, H., and D. M. Scott. 2008. An integrated spatio-temporal GIS toolkit for exploring intra-household interactions. *Transportation* 35:253–68.

Kanungo, T., D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu. 2002. An efficient *k*-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (7): 881–92.

Kitamura, R. 1988. An evaluation of activity-based travel analysis. *Transportation* 15:9–34.

Koppelman, F. S., and E. I. Pas. 1985. Travel-activity behavior in time and space: Methods for representation and analysis. In *Measuring the unmeasurable*, ed. P. Nijkamp, H. Leitner, and N. Wrigley, 587–627. Dordrecht, The Netherlands: Martinus Nijhoff.

Kwan, M.-P. 1998. Space–time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis* 30 (3): 191–216.

———. 1999. Gender, the home–work link, and space–time patterns of non-employment activities. *Economic Geography* 75 (4): 370–94.

———. 2000. Interactive geovisualization of activity travel patterns using three-dimensional geographical information systems: A methodological exploration with a large dataset. *Transportation Research Part C* 8 (1): 185–203.

———. 2004. GIS methods in time-geographic research: Geocomputation and geovisualization of human activity patterns. *Geografiska Annaler B* 86 (4): 267–80.

———. 2008. From oral histories to visual narratives: Representing the post-September 11 experiences of the Muslim women in the USA. *Social and Cultural Geography* 9 (6): 653–69.

———. 2013. Beyond space (as we knew it): Toward temporally integrated geographies of segregation, health, and accessibility: Space–time integration in geography and GIScience. *Annals of the Association of American Geographers* 103 (5): 1078–86.

Kwan, M.-P., N. Xiao, and G. Ding. 2014. Assessing activity pattern similarity with multidimensional sequence alignment based on a multiobjective optimization evolutionary algorithm. *Geographical Analysis* 46 (3): 297–320.

Laube, P. 2014. *Computational movement analysis*. New York: Springer.

Laube, P., S. Imfeld, and R. Wiebel. 2005. Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographic Information Science* 19:639–68.

Lee, J. Y., and M.-P. Kwan. 2011. Visualisation of socio-spatial isolation based on human activity patterns and social networks in space–time. *Tijdschrift voor Economische en Sociale Geografie* 102 (4): 468–85.

Lew, A., and B. McKercher. 2006. Modeling tourist movements: A local destination analysis. *Annals of Tourism Research* 33 (2): 403–23.

Long, J. A., and T. A. Nelson. 2013. A review of quantitative methods for movement data. *International Journal of Geographical Information Science* 27 (2): 292–318.

MacPherson, D. W., and B. D. Gushulak. 2001. Human mobility and population health: New approaches in a globalizing world. *Perspectives in Biology and Medicine* 44 (3): 390–401.

Matthews, S. A., and T.-C. Yang. 2013. Spatial polygamy and contextual exposures (SPACEs): Promoting activity space approaches in research on place and health. *American Behavioral Scientist* 57 (8): 1057–81.

Meratnia, N., and R. A. de By. 2002. Aggregation and comparison of trajectories. In *Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems*, 49–54. New York: ACM

Murdock, V. 2011. Your mileage may vary: On the limits of social media. *SIGSPATIAL Special* 3 (2): 62–66.

National Research Council (NRC). 2002. *Community and quality of life: Data needs for informed decision making*. Washington, DC: Board on Earth Sciences and Resources, Division on Earth and Life Studies, National Research Council.

Neutens, T., F. Witlox, N. van de Weghe, and P. de Maeyer. 2007. Human interaction spaces under uncertainty. *Transportation Research Record* 2021:28–35.

Openshaw, S., and P. J. Taylor. 1981. The modifiable areal unit problem. In *Quantitative geography: A British view*, ed. N. Wrigley and R. J. Bennett 60–70. London and New York: Routledge and Kegan Paul.

Palmer, J. R. B., T. J. Espenshade, F. Bartumeus, C. Y. Chung, N. E. Ozgencil, and K. Li. 2013. New approaches to human mobility: Using mobile phones for demographic research. *Demography* 50:1105–28.

Pas, E. I. 1988. Weekly travel-activity behavior. *Transportation* 15:89–109.

Pas, E. I., and F. S. Koppelman. 1987. An examination of the determinants of day-to-day variability in individuals' urban travel behavior. *Transportation* 14:3–20.

Purves, R. S., R. Laube, M. Buchin, and B. Speckmann. 2014. Moving beyond the point: An agenda for research in movement analysis with real data. *Computers, Environment and Urban Systems* 47 (2014): 1–4.

Ren, F., and M.-P. Kwan. 2007. Geovisualization of human hybrid activity travel patterns. *Transactions in GIS* 11 (5): 721–44.

Shaw, S. L., and D. Wang. 2000. Handling disaggregate spatiotemporal travel data in GIS. *GeoInformatica* 4 (2): 161–78.

Shaw, S. L., and H. Yu. 2009. A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *Journal of Transport Geography* 17 (2): 141–49.

Shaw, S. L., H. Yu, and L. S. Bombom. 2008. A space–time GIS approach to exploring large individual-based spatiotemporal datasets. *Transactions in GIS* 12 (4): 425–41.

Shen, Y., M.-P. Kwan, and Y. Chai. 2013. Investigating commuting flexibility with GPS data and 3D geovisualization: A case study of Beijing, China. *Journal of Transport Geography* 32:1–11.

Silm, S., and R. Ahas. 2014. Ethnic differences in activity spaces: A study of out-of-home nonemployment activities with mobile phone data. *Annals of the Association of American Geographers* 104 (3): 542–59.

Sinha, G., and D. M. Mark. 2005. Measuring similarity between geospatial lifelines in studies of environmental health. *Journal of Geographical Systems* 7:115–36.

Van der Knaap, W. G. 1999. Research report: GIS-oriented analysis of tourist time–space patterns to support sustainable tourism development. *Tourism Geographies* 1 (1): 56–69.

Wang, D., and B. Sun. 2015. Detecting activity types and trip purposes from passive GPS data: A data mining approach. In *Space–time integration in geography and GIScience: Research frontiers in the US and China*, ed. M.-P. Kwan, D. Richardson, D. Wang, and C. Zhou, 211–34. Dordrecht, The Netherlands: Springer.

Winter, S. 2009. Towards a probabilistic time geography. In *Proceedings of the 17th ACMSIG SPATIAL international conference on advances in geographic information systems*, ed. M. Mokbel, P. Scheuermann, and W. G. Aref, 528–31. New York: ACM.

Winter, S., and Z. C. Yin. 2010. Directed movements in probabilistic time geography. *International Journal of Geographical Information Science* 24 (9): 1349–65.

Wong, D. W. 2009. The modifiable areal unit problem (MAUP). In *The Sage handbook of spatial analysis*, ed. A. S. Fotheringham and P. A. Rogerson, 105–23. London: Sage.

Yu, H. 2007. Visualizing and analyzing activities in an integrated space–time environment: Temporal geographic information system design and implementation. *Transportation Research Record: Journal of the Transportation Research Board* 2024:54–62.

Yuan, M., and A. Nara. 2014. Space–time analytics of tracks for understanding of patterns of life. In *Space-time integration in geography and GIScience: Research frontiers in the US and China*, ed. M.-P. Kwan, D. Richardson, D. Wang, and C. Zhou, 373–98. Dordrecht, The Netherlands: Springer.

Correspondence: Department of Geography, University of Wisconsin–Madison, Madison, WI 53706, e-mail: qhuang46@wisc.edu (Huang); Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA 22030, e-mail: dwong2@gmu.edu (Wong).