Dynamic diurnal social taxonomy of urban environments using data from a geocoded time use activity-travel diary and point-based business establishment inventory

Seo Youn Yoon a,⇑, Srinath K. Ravulaparthy b,1, Konstadinos G. Goulias b,1

a Korea Research Institute for Human Settlements, 254 Simin-daero, Dongan-Gu, Anyang-Si, Gyeonggi-do 431-712, South Korea
b GeoTrans Lab., Department of Geography, University of California Santa Barbara, Santa Barbara, CA 93106, USA

Article info
Article history:
Available online 11 February 2014

Keywords:
Social interaction
Location choice
Dynamic place characteristics
Social taxonomy of urban environments
Human–environment relationships
Spatio-temporal kriging

Abstract
In this paper, we explore the diurnal dynamics of joint activity participation in a small city in Pennsylvania, USA, using behavioral data and an inventory of business establishments. We account for the variation caused by the collective impact of social, temporal and spatial choices of individuals to produce predicted space–time visualizations of activity participation. The focus is on how social contexts of an activity impact the temporal and spatial decisions regarding the activity locations and how this impact varies depending on activity types. A comparison across activity types and social interaction types is made among spatial patterns during a day. The CentreSIM dataset, which is a household-based activity diary survey collected in Centre County (Pennsylvania, USA) in 2003, provides very detailed social interaction information enabling the analysis of social, spatial and temporal aspects of activity participation. In this paper we use this information to develop a spatio-temporal interpolation method and demonstration based on kriging. In this way, we extract the dynamic social taxonomy of places from the behavioral information in the dataset and suggest how urban and transportation models can be informed from the dynamics of places by observing “what is taking place” (activities being pursued in the context of this paper) combined with “what exists” (business establishments) or “what is available” (businesses that are open). The method here can also be used to improve the design of urban environments (e.g., filling gaps in desired activity locations), manage specific places (e.g., extending the opening and closing times of businesses), study transportation policies that are sensitive to time of day (e.g., pricing of parking to discourage crowding and traffic congestion), and modeling of spatio-temporal decisions of social activities in travel demand models (e.g., to guide the development of model specification and representation of the space in which behavioral models are applied).

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Urban environments transform in cyclical manner offering opportunities for activity engagement that change by time of day, day of week, month of year and so on. This variability results in multiple layers of dynamics that should be included in behavioral data collection (Goulias et al., 2013) and in activity-based travel demand simulators (Kitamura et al., 1997;
Activity participation, time allocation, and travel are all heavily influenced by human interactions and early activity-based methods recognized their importance (Kitamura, 1988; Jones, 1990; Bhat and Koppelman, 1999). Although these interactions motivate the collection of behavioral data using household surveys and the creation of models that aim at replicating household activity schedules (Bhat et al., 2013), activity participation and travel of people takes place in many social fields (or networks) that involve many persons outside the household that exert social influence (Bhat and Pendyala, 2005; Dugundji and Walker, 2005; Carrasco and Miller, 2006; Carrasco et al., 2008; Habib et al., 2008; Farber and Páez, 2009; Arentze and Timmermans, 2008; Timmermans and Zhang, 2009; Assen, 2010; Goulias and Yoon, 2011; Walker et al., 2011; Deutsch and Goulias, 2013). This social influence and interactions are also place specific and it is informative to examine them employing the time–space prism (Schönfelder, 2006; Neutens et al., 2008; Yoon and Goulias, 2010).

In spite of all these advancements, a gap is left in the development and testing of a method that produces time varying spatial distributions of activity participation propensity that can be used for multiple purposes. Developing these dynamic spatial distributions is our ultimate objective in this research and this paper is the first attempt. Our motivation also stems from the following simple issue. Neither behavioral data nor business establishment databases can answer the question of what will a person do at a specific location at a specific time of day with high spatial and temporal resolution. Even when one employs sophisticated urban simulators the activity location is either a traffic analysis zone (i.e., comparable to a large city block) or some other kind of spatial summary of points. For this reason post-processing of simulated data is required to position (geolocate) each activity and each person at a specific business establishment and a specific period of a day (Tang et al., 2013). Instead of ad hoc geolocation algorithms we could use a spatio-temporal distribution. Developing this type of distribution for activity participation requires first understanding how urban space is used and possibly interpolate data in a way that produces a high resolution distribution of the propensity to participate in activities by the residents and visitors of a place. It is also important to use these distributions to identify urban policies that will guide people to behave in an environmentally responsible manner and to build simulation models that create policy scenarios.

There are some complications as we attempt to do all this. One of the complicating issues is due to the standardized system of business classification (International Harmonized and SIC/NAICS see Pierce and Schott, 2012), which is not sufficient for understanding how each business establishment is used. For example, consider a business establishment classified as restaurant. A person may go there to have a breakfast meeting with a business customer at 8:00 am, have lunch with his co-workers at 12:00 noon, meet with a friend for coffee at 3:00 pm, have dinner with his family at 6:00 pm and come back at 11:00 pm to listen to live music and have a drink. This restaurant changes in terms of the activity engaged and the interaction of the people involved. In response to the opportunities offered by built environments, individuals and their groups choose to populate certain places that satisfy their need to pursue certain activities at specific times in a day, and engage in activities alone or with other people. This choice is an outcome of many other events and offers regularity that we can exploit in spatial prediction. Spatio-temporal patterns of activity engagement are constrained by the dynamic patterns of attraction that are a function of the demand for services but also dictated by regulations as well as the availability of other persons that are involved in the activity. To develop the method in this paper we follow a vein of development that started 40 years ago.

Behavioral dynamics were discussed as one of the important aspects of urban environments in “what time is this place?” by Lynch (1972). Lynch stated that human activities are as important as or even of greater importance than the permanent physical artifacts to the quality of a place. Time of day and timing are important ingredients in understanding the role places play in the life of people and the places themselves. In fact, for human behavior the temporal aspects are as important as the spatial aspects. This is also a major focus of Chapin’s analysis (1974), and received attention as early as the first georeferenced time-use study in Halifax was made available (Harvey, 1993; Janelle and Goodchild, 1983; Goodchild et al., 1993). Spatio-temporal patterns of activities in Halifax in these studies are shown as the presence of people at different parts of the city at different times in a day, and these same dynamic patterns are analyzed for different socio-demographic groups. These daily dynamics of a place have been represented mostly as variability of available opportunities by considering opening and closing hours of businesses (Weber and Kwan, 2002; Kim and Kwan, 2003; Chen et al., 2011; Yoon et al., 2012). All these models focus on what exists and what is available at a place at a given time point with personal trajectories added within the Hagerstrand time–space prism and related aquarium. In this paper, we combine the ideas from the Halifax study with the focus more on the temporal patterns of activities taking place to describe the dynamic change of a place especially as it relates to social interaction, adding information about social networks accessing each location by time of day. With inspiration from the probability fields in Beckmann et al. (1983a,b), we derive, using data, time varying spatial distributions of activities that are informed by social influence, capture the dynamics of activities focusing on social interaction taking place.
at each location. Also, location choices and timing of activities are closely tied to the type of social interaction involved in the activities (i.e., locational and temporal choices for eating out with children and eating out with friends are different for the parents but sometimes occurring at the same location) and this information on the joint patterns is useful in expanding models of activity participation agendas, activity scheduling, and destination choice.

In the next section the data used here are described, followed by two distinct patterns of time allocation at home and outside the home. This is followed by the spatio-temporal interpolation with kriging formulation that includes time as an added dimension. The paper concludes with a summary, findings, and next steps.

2. Data used

The CentreSIM dataset is a household-based activity diary survey collected in Centre County, Pennsylvania, USA in 2003. Each person in the household provided a 2-day complete record of in-home and out-of-home activities including detailed information on with whom and for whom the activities were pursued (Patten and Goulias, 2004; Goulias and Kim, 2005; Goulias and Henson, 2006; Goulias and Yoon, 2011). After data cleaning and verification 1471 persons (from 718 households), which is approximately 1% sample of Centre County population, were selected for the analysis in this paper.

Respondents were asked with whom they participated in their activities and how many people participated in the activities. The persons who joined the activities were classified depending on their relationship with the respondent, and the relationship with the respondent was described using up to five “with whom” variables from the diary. Of note, when the respondent participated in an activity, for example, with her husband, and two daughters, their relationship can be coded as “multiple family members” in the first variable with the other four “with whom” variables left null. Out of the total 34,635 activity episodes, the five “with whom” variables have 16,984, 2447, 425, 64, and 11 not null values respectively. The types of social contacts are listed in Table 1. Among those joint activities, 11,812 episodes were only with family members (9914 at home activities, 1818 out of home activities including being at other family member’s place) and 1092 episodes were only with friends (369 at home activities, 723 out of home activities including being at friend’s place).

The activities with family are mostly reported by all the individuals who participated in the activity except for extended family members. For example, when a family of three members had lunch at a restaurant, the activity was reported three times, once by each household member in the survey. However, the activities with friends are reported only once by the respondents and the number of friends who joined the activities is reported as a separate variable.

The spatio-temporal diagrams (called 3D aquarium herein following the Hagerstrand term used in Kwan, 1999b, 2000, 2004) of Fig. 1 show how differently the study area is populated by different types of social interaction at each time in a day. The vertical axis represents time of day starting from midnight at the lowest point to the midnight of 24 h later at the highest point, and the colors show the starting time of each activity. For example, the place marked with a square, which is the portion of the town called State College and adjacent to the Pennsylvania State University campus, which is called University Park (UP). State College in its adjacent city block to UP contains restaurants, bars and many other businesses such as retail clothing stores. It is relatively densely frequented with solo activities including work and school activities during the day, becomes a family place from 2 pm to 6 pm, and transforms into a place for interaction with friends after 8 pm. The social activities with friends continue until 2 am (see short activities after midnight in Fig. 1-C). On the other hand, the spatial distribution of social activities with family appears more dispersed in space than that of social activities with friends.

From the exploratory plotting of activities, quite different spatio-temporal patterns are observed for activities with different social groups. In the next section, temporal aspects of social activities are discussed in more detail. Then, spatio-temporal interpolation is conducted for out-of-home eating and shopping activities with family and friends (four activity types), while accounting for the business locations (i.e., business establishments) that correspond to eating and shopping activities.

3. Analysis of within-day temporal patterns

3.1. In-home social activities

Fig. 2 shows the temporal distribution of in-home activities that involve family or friends. For activities with family (Fig. 2–a), a relatively small number of activities are reported from 8 am to 6 pm compared to the other times of day, and

<table>
<thead>
<tr>
<th>With whom 1</th>
<th>With whom 2</th>
<th>With whom 3</th>
<th>With whom 4</th>
<th>With whom 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td>12,441</td>
<td>1620</td>
<td>304</td>
<td>33</td>
</tr>
<tr>
<td>Friends</td>
<td>1172</td>
<td>215</td>
<td>36</td>
<td>5</td>
</tr>
<tr>
<td>Roommates</td>
<td>100</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Work-related</td>
<td>1763</td>
<td>156</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>School-related</td>
<td>572</td>
<td>113</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Other people</td>
<td>760</td>
<td>294</td>
<td>71</td>
<td>21</td>
</tr>
<tr>
<td>Pets and animals</td>
<td>176</td>
<td>44</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>16,984</td>
<td>2447</td>
<td>425</td>
<td>64</td>
</tr>
</tbody>
</table>

Table 1 Social activities with different social groups.
there is only a small peak of eating activities with family at lunch time. After the work hours, there is the highest peak of eating activities at 6 pm, and the number of activities for watching movies, reading books, and using computer increases until 10 pm. This is a remarkable regularity that our sample exhibits. Overall, this temporal distribution of the in-home activities with family shows a very clear impact of out-of-home work and school hours between 8 am and 6 pm.

On the other hand, the interaction with friends (Fig. 2-b) is not affected by out-of-home work hours as clearly as the interaction with family presumably due to the different socio-demographic characteristics of the respondents who reported interaction with family and the respondents who reported interaction with friends. The eating activities with friends extend until later hours than the eating activities with family because eating activities serve various purposes when different social interactions are involved. In addition, watching movies, reading books, using computer, playing games and socializing extend until late at night (2 am) when they are with friends. Sociality with friends exhibits a different time signature than with family presumably reflecting different types of binding constraints. For example, the regularity of school going children imposes stricter constraints on parents to allow them time after dinner to study and sleeping time to be ready for school the next day. Adults with no children may not face the same temporal constraints.

These patterns show that the temporal decisions on the same activity type taking place at home can be different depending on the social context. The temporal decisions include beginning and ending times, duration, arrival at and departure from home to pursue the in-home social activities. In addition, the interaction with out-of-home activities (i.e., bargaining and compensation in time use, sequencing) will also be different depending on the social context of the in-home social activities.

### 3.2. Out-of-home social activities

In this section, temporal patterns of out-of-home social activities are analyzed, but the analysis is limited only to the activity types that can directly be associated with specific types of businesses and were reported for a sufficient number of times in the survey to develop distributions. One of the activity types selected is EATING activities that includes eating meals at restaurants and fast food restaurants, drinking coffee, eating ice creams and desserts, and drinking at bars. This
covers all the consumption of food and beverages reported in the survey. The other activity type is SHOPPING activities including shopping at grocery stores, convenience stores, office and home supply stores, pharmacies, electronics stores, big retail stores (Walmart, Target, etc.), and big malls, and shopping for clothes, shoes and anything to wear as well. However, shopping activity excludes picking up food from restaurants and fast food restaurants.

In the four quadrants of Fig. 3, the time-of-day distribution of each activity type with family and with friends is shown for weekdays and weekends in black and white respectively. Comparison between eating activities with family and friends shows that eating with family is very concentrated at regular eating hours (noon to 1 pm for lunch and 6–8 pm for dinner) and eating with friends is more spread throughout the day. Eating with family on weekends shows similar temporal distribution, but tends to start and end earlier than on weekdays, and eating with friends on weekends are pursued more in the evening than for lunch and tends to end earlier (no activities around 1 am on weekends) than on weekdays. Based on these observations, during lunch time in weekends, restaurants in the study area are more likely to be populated by groups of families than by groups of friends. On the other hand, shopping on weekdays shows a clearer peak in the evening when it is with friends than when it is with family. On weekends, however, shopping with friends is spread throughout the day. Shopping with family on weekdays is spread from 9 am to 9 pm without any clear peaks, but shows a very clear peak between noon and 3 pm on weekends. The temporal profiles of eating and shopping activities show that these activities are pursued at different times of day depending on which social interaction is involved with the activities. Each of these two activity types has its own temporal pattern in the distribution in addition to the spatial correlation pattern that was briefly observed in the spatio-temporal plots of Fig. 1. In the next section, we propose a spatio-temporal interpolation method that is able to reflect

![Fig. 2. In-home activities with family and friends.](image-url)
the activity-type-specific spatio-temporal correlation patterns to find diurnal dynamics of social activities at different places based on this sparse sample of observations.

4. Spatio-temporal interpolation

4.1. Conceptual design of spatio-temporal interpolation

The spatio-temporal interpolation method used in this paper is based on an expanded form of geo-statistical interpolation method. In many cases, kernel density has been used to generate density of activity locations for individuals or a probabilistic field of activity locations based on sample observations (i.e., Schönfelder and Axhausen, 2002; Kwan, 2000). The difference between spatio-temporal interpolation used in this study and kernel density are:

- spatio-temporal interpolation considers both spatial distribution and time of day whereas kernel density has been used for the spatial dimension in most cases (although it is possible to consider space and time at the same time using kernel density as well), and
- spatio-temporal interpolation uses spatio-temporal correlation patterns that are specific for each activity type to generate an activity-specific probabilistic field, which is not possible when kernel density is used.

Since the survey data contain approximately 1% of the population, we assume that the observed activities capture the dynamic patterns of the activities of the population. One could also develop other techniques that use sample weights or
population synthesis that are not our concern in this paper and will be discussed in the conclusions. Inevitably we miss certain spots in the spatio-temporal aquarium of a day where (and when) activities are actually being pursued by the population. On the other hand, there are locations where activities cannot be pursued at all because there are no opportunities offered at those locations. To capture all this, the spatio-temporal aquarium of the study area is divided into 250 m by 250 m by 1 h cells, and the cells are classified into three categories:

- cells with observed activities (gray cells with observed activity intensity in Fig. 4),
- cells with no opportunities (white cells), and
- cells offering opportunities but with unknown activity intensity (no observation in the sample, cells shaded with lines).

The number of people who participated in a certain type of activity within a spatio-temporal cell is considered as the observed intensity of the activity type at a location during each 1-h period. The number of people who joined respondents for an activity is also included in the computation of observed intensity.

The rationale of using spatio-temporal correlation to fill in the values of the cells with no observation can be described as the following, and the illustration of Fig. 4 is provided to assist understanding. The circled numbers at the leftmost corner of each cell is the number assigned to the cell. The dark number in the middle of each cell is the intensity of activities (number of people observed).

- Each type of social activities has a certain level of spatial and temporal correlation in terms of the intensity of the activities.
- Therefore, the intensity of an unknown cell, for example cell number 4 in Fig. 4, has a higher correlation with the observations in the adjacent/bordering cells 3, 8, and 5 (intensity of 8, 10, and 6 respectively) than another unknown cell, cell 6, does, because cell 6 is located farther away from cells 3, 8, and 5 than cell 4 is.
- When a certain type of activity is observed at 6 pm and 8 pm in a cell, it is very likely that there are activities of the same type at 7 pm as well due to temporal correlation (see the relationships among cells 1, 2, and 3).

As the first step of spatio-temporal interpolation, spatial and temporal correlograms are computed and estimated.

### 4.2. Variogram

Let \( Z(s) \) represent the stochastic process or a spatial random field under consideration. In this case, the spatial autocorrelation or spatial dependence between two pairs of observations \( Z(s_i) \) and \( Z(s_j) \) for locations \( s_i \) and \( s_j \) grouped by a separation distance \( h_{ij} = |s_i - s_j| \) is given by the function \( \gamma(h) \) which is called the semivariogram or variogram of the random field \( Z(s) \). In our analysis, the random field \( Z(s) \) is the spatial distribution of activity intensity under consideration. Therefore, semivariogram is expressed as the expected squared increment of values between locations \( s_i \) and \( s_j \) as indicated by Wackernagel (2003).

\[
\gamma(h) = \frac{1}{2} E[(Z(s) - Z(s + h))^2]
\]

Under stationarity assumptions for process \( Z(s) \) we can state that the variance of \( Z \) is constant and spatial correlation of \( Z \) does not depend on location \( s \), but only on separation distance \( h \). The semivariogram of a spatial random field \( Z \) describes the spatial structure of the phenomenon under investigation. This spatial structure is characterized by three different components of the semivariogram: (a) sill is defined as the semivariance value at the point where the variogram levels off; (b) range is defined as the lag distance at which the semivariogram reaches the sill value (autocorrelation is essentially zero in the range).

![Fig. 4. Conceptual design of spatio-temporal interpolation.](image-url)
beyond the range); and (c) nugget represents the variability at distances smaller than the typical sample spacing, including the measurement error.

The variogram is often used for spatial prediction or interpolation or simulation of the observed process based on point observations. To ensure that the predictions are associated with non-negative predictions of the variances, the matrix with semivariance values between observation points and any possible prediction point need to be non-negative definite. One common way is to infer parametric variogram models from the data. The traditional way of finding the suitable variogram model is to fit a parametric model to the sample variogram defined in Eq. (1). The parameters that define the fitted variogram model are partial sill, range and nugget.

The most common models used in geo-statistics are exponential, spherical, and Gaussian models with or without a nugget. In this paper, we adopt the exponential model for parametric variogram model for all activities with both family and friends. For example, for shopping activities with family, the exponential model has the variogram model parameters with partial sill as 0.05, range as 1164.786 m and a nugget of 0.035. The model parameters were chosen based on the weighted least squares fitting of the variogram model to the sample variogram model as proposed by Cressie (1990). As described by Bivand et al. (2008), the fitting method uses non-linear regression to find the coefficients with a weighted sum of squared errors minimized to fit the parametric model. The spatial correlograms $\rho(h)$ for the reported shopping and eating activities with family and friends are also computed using the following equation:

$$\rho(h) = \frac{S - \gamma(h)}{S}$$

where $S$ is the partial sill as obtained from the fitted variogram model and $\gamma(h)$ is the semivariance from the variogram with a lag distance of $h$.

In a similar way only replacing distance with temporal difference, the temporal correlograms for the reported activities are computed. The temporal correlogram is the autocorrelation function for time lag $t$. The sample spatial and temporal correlograms for family shopping are as illustrated in Fig. 5.

4.3. Kriging

As noted by Goovaerts (1997) geo-statistics is also commonly applied to natural and social sciences. Kriging is a geo-statistical technique which is a type of interpolation based on regression against an observed set of values from the process $Z(s)$, weighted according to spatial covariance values (Bailey and Gatrell, 1995). Kriging predicts unknown values from observed data at known locations. This method uses variogram to express the spatial variation minimizing the error of predicted values. As noted by Bailey and Gatrell (1995), kriging allows deriving weights that result in optimal and unbiased estimates.

Within a probabilistic framework, kriging attempts to (a) minimize the error variance and (b) systematically set the mean of the prediction errors to zero, so that there are no over – or underestimates. There are different kinds of kriging such as simple kriging, universal kriging and cokriging (Isaaks and Srivastava, 1989). Applications of kriging in transportation engineering and planning includes identification of mismatches between transit service and employment centers (Modarres, 2003), interpolation of Average Annual Daily Traffic from traffic counts (Shamo et al., 2012; Wang and Kockelman, 2009;
Selby and Kockelman, 2013), highway section travel times and average speeds (Aultman-Hall and Du, 2006; Miura, 2010; Ozbay and Yildirimoglu, 2011; Zou et al., 2012) as well as generalized model building tutorial on non-stationary spatial models (Vichiensan et al., 2006). In addition, Pitombo et al. (2010) specifically investigates the applicability of kriging with external drift (or universal kriging) for forecasting urban trip generation for the metropolitan area of Sao Paulo, Brazil. To the best of our knowledge and at the time this paper was written no application of kriging exists in activity-based spatio-temporal models (Phillips and Marks, 1996; Isaaks and Srivastava, 1989).

As described previously let \( Z(s) \) represent the spatial random field that is composed of mean \( m(s) \) and residual \( e(s) \) as represented in (3). The mean \( m(s) \) varies spatially and can be modeled as a linear function of known predictors \( X(s) \) and to be estimated regression coefficients \( \beta_j \) as shown in the following equation:

\[
Z(s) = m(s) + e(s)
\]

\[
Z(s) = \sum_{j=0}^{p} X_j(s) \beta_j + e(s) = X\beta + e(s)
\]

4.4. Ordinary Kriging

X\( _j(s) \) forms each column of the \( n \times (p + 1) \) design matrix, which is composed of the \( p \) predictors or explanatory variables for each of the \( n \) observations. \( \beta \) is the column vector with \( p + 1 \) unknown coefficients, which also includes an intercept to be multiplied with \( X_0(s) = 1 \). When the number of predictors (\( p \)) equals to zero, which means the equation includes no explanatory variables but the intercept only, the corresponding prediction is called ordinary kriging. Ordinary kriging is mathematically equivalent to random field \( e(s) \). This spatial random field process \( Z(s) \) has a constant mean \( m \), which is not dependent on any location \( s \). In other words, \( m(s) = m \). By using a weighted linear combination of observed values \( z_i \) at the sample locations, the prediction for location \( s \) is given as

\[
\hat{Z}(s) = \sum_{i=1}^{n} w_i(s)Z(s_i)
\]

The values of \( w_i(s) \) are chosen so that the mean value of the predicted random field \( Z(s) \) is constrained to be \( m \) and the weights \( w(s) \) sum to one. In this study, the spatial prediction or the kriging variable is the number of people participating in activities for eating and shopping with family and friends respectively. Furthermore, associated with the kriging predictions are also the error variance or standard deviations of the attribute of interest at each location in this analysis. Therefore, the uncertainty of the interpolation or kriging surface can be quantified from the error variance (or residuals) using variety of techniques such as confidence intervals of the point estimates (or predictions), bootstrapping and Monte-Carlo simulations (Phillips and Marks, 1996; Isaaks and Srivastava, 1989).

4.5. Space–time kriging

Space–time kriging is a well-demonstrated interpolation methodology that considers spatial and temporal correlations simultaneously as introduced by De Cesare et al. (2001). It is fundamentally the same as the traditional spatial kriging except in the handling of the covariance modeling (To and Maidment, 2009). As indicated by To and Maidment (2009) and Pebesma (2012), we can treat time as an added dimension in spatial kriging and then add the lag distances along spatial and temporal axes as vectors to get a single Euclidian space–time metric (spatio-temporal distance). In this case, the semivariance between locations \( s_i \) and \( s_j \) given by \( \gamma(s_i, s_j) \) is a function of the Euclidian space–time metric as represented in the following equation:

\[
\gamma(s_i, s_j) = f(h, t)
\]

where \( h \) is the spatial lag between location \( s_i \) and \( s_j \) and \( t \) is the time lag between time periods \( t_1 \) and \( t_2 \).

We use space–time kriging to interpolate the intensity of activities (number of persons) observed from the sample, and the interpolation accounts for both spatial and temporal dependencies. As identified by Pebesma and Wesseling (1998) kriging can be extended from two dimensions to three dimensions. When the spatial domain is constrained to two dimensions, the third dimension can be used to represent time. In such cases, the space–time metric variogram model, which allows for geometric anisotropy definition in three dimensions, can be used for space–time kriging. However, when defining the three-dimensional variogram model, one can use anisotropy coefficients between space–time axes using the lags where the correlations reach the same level (i.e., zero correlation) to approximate the space–time variogram model (Pebesma, 2012).

Thus to achieve the objective of space–time kriging, we estimate both the spatial and temporal correlograms of the number of people participating in activities for eating and shopping with family and friends. For example, in the case of temporal correlation for family shopping activities estimated based on the sample temporal correlation shown in Fig. 5 the zero correlation occurs at a time lag of approximately 14 h. Similarly, in the case of spatial correlation for family shopping activities estimated based on the sample spatial correlation also shown in Fig. 5 the zero correlation occurs at a distance of approximately 1165 m. This zero correlation occurrence in space and time is now converted into a single Euclidian space–time metric by defining the anisotropy coefficient as 14 h/1165 m. In this way, the space–time variogram model described above is applied for the space–time kriging of activity intensity.
4.6. Steps for space–time kriging

Based on the discussion in the previous section following are the steps for implementing space–time kriging for the number of people participating in eating and shopping activities with family and friends. We used the R library package "gstat" to perform ordinary kriging. Refer to Pebesma (2001) for more information on this package.

**Step 1:** Compute the empirical (i.e., sample) correlograms for the number of people participating in activities.

![Figure 6. Comparison between observation and prediction. Each dot: activity intensity measured or estimated as number of people within a 250 m by 250 m cell. Red and large dots: high intensity, green and small dots: low intensity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image)

![Figure 7. Predicted activity intensity of eating activities with family and eating activities with friends on weekdays (green and small dots: low intensity, yellow and medium dots: medium intensity, and red and large dots: high intensity). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image)
Step 2: Compute the empirical variograms for the number of people participating in activities.

Step 3: Fit a parametric variogram model to the empirical variogram. In this study an exponential variogram model is fitted with partial sill, range and nugget as the parameters.

Step 4: Construct the single Euclidian space–time metric by defining the anisotropy coefficient. The anisotropy coefficient is the inverse of the distance lag at which the spatial correlation is zero, which is equivalent to the time lag at which the temporal correlation is zero.

Step 5: Define the space–time variogram model with all the relevant parameters of partial sill, range, nugget and anisotropy coefficient as obtained from Step 3 and Step 4.

Step 6: Perform ordinary kriging for the number of people participating in activities with the defined space–time variogram as in Step 5.

5. Analysis of the interpolation results

We conducted spatio-temporal interpolation within a spatio-temporal aquarium of 50 cells (12.5 km) by 39 cells (9.75 km) by 24 h for the four types of social activities on weekdays respectively. The area includes portion of State College and the campus of Penn State University, which is located approximately at the center of the following maps.

Fig. 6 shows how the interpolation fills in non-observation cells using the spatio-temporal correlation for the case of eating-out with friends. First of all, we can see that not only the spatial gaps are filled but also the observations of each time slice are used to fill the non-observation cells of the other time slices. Second, the existence of activity clusters, especially the area marked with a circle, is much clearer in the prediction maps than in the observation maps. Third, the temporal variation of activity intensity in the clusters is much more noticeable in the prediction maps than in the observation maps. The strongest cluster in the circle is actually co-located with the higher density business (restaurants, fast food restaurants, coffee shops and night club/bars in this case) area adjacent to the PSU campus, which is an informal way to show that this methodology is able to find out association of a business cluster with specific social interaction types at specific times of a day and to provide information that defines social places. This cannot be done using business establishment information alone.

Further interpretation of the association between places and social interaction types is given based on the comparison between eating-out with family and eating-out with friends (Fig. 7). The area indicated with a dashed circle is another cluster.
of businesses, but this is a typical suburban shopping mall with big box stores. The prediction shows that restaurants at this place are more populated by persons with family than persons with friends between 6 pm and 8 pm. On the other hand, the place indicated with a solid circle is populated by people with family only around 6 pm and far less populated by family after 7 pm than the place of the dashed circle is. In addition, the temporal patterns of the two eating activities at different places are also notable. The eating activities with family fade out after 9 pm in the dashed circle and after 7 pm in the solid circle. On the other hand, the eating activities with friends fade out after 7 pm in the dashed circle, but persist until 1 am in the solid circle, which again characterize each place. This output effectively shows how places change in their attractiveness for different activities and personal contacts by time of day and how different types of social interactions happen at the same place but at different times.

A 3D visualization of the observation and the prediction for each activity type is provided in Fig. 8 enabling overall comparison of spatio-temporal distributions across activity types. The prediction shows continuous distribution with a clear indication of spatio-temporal activity clusters and temporal variation of activity intensity at different places.

6. Summary and conclusions

In this paper we demonstrate the use of a spatio-temporal interpolation method based on kriging. We extract the dynamic social taxonomy of places from the behavioral information in a time use activity-travel diary and show how urban and transportation models can be informed from the dynamics of places by observing “what is taking place” (activities being pursued in the context of this paper) combined with “what exists” (business establishments) or “what is available” (businesses that are open). As one would expect social context is a very important determinant in spatio-temporal decision making for activities echoing other analyses that examine time alone (Habib and Carrasco, 2011) and this collectively contributes to the signature characteristics of places by time of day. From this relationship we can map how places attract different social interactions at different times in a day. This dynamic aspect stresses the importance of obtaining information on when activity opportunities are offered by businesses and merge this information with when actual activities are conducted and with whom. The results here can complement other available data that are used for activity-based travel demand models and guide new data collection. In fact, the analysis here is another proof that precise geocoding all activities in a diary is needed and collecting data about business establishments at the same period during which a household-based activity and/or travel diary is collected enables the creation of dynamic social taxonomies of places. The method here can also be used to improve the design of urban environments (e.g., filling gaps in desired activity locations), manage specific places (e.g., extending the opening and closing times of businesses), study transportation policies that are sensitive to time of day (e.g., pricing of parking to discourage crowding and traffic congestion), and modeling of spatio-temporal decisions of social activities in travel demand models (e.g., to guide the development of model specification and representation of the space in which behavioral models are applied).

This is a first attempt to study the feasibility in using a sparse database of activity participation to aid in spatio-temporal interpolation and in essence impute data in space and time. The activity episodes used, however, were considered as events that are independent from other activities of a person and her household in the data (except for the fact that we distinguish between the two group of family and friends). In fact, every meal a person takes is within a daily schedule of activities that are interdependent in space and time. When this person visits a restaurant with her family, this activity is also part of the schedule of each member of her family. Ideally, one should be developing a kriging interpolation that takes into account the interaction among different spatial distributions and/or find other ways to capture these interactions at least at the day level (e.g., using a synthetic schedule estimator of the SimAGENT type see Bhat et al., 2013).

Recall that the spatial distribution of activity opportunities was used to classify cells in the kriging support that have business establishments and differentiate them from cells that have no business opportunities to avoid predicting activities in places that do not support them. This is a simplified way of representing spatial opportunities and there are better and more informative ways (e.g., through time varying accessibility as in Chen et al., 2011). Possible extension of the work here includes inclusion of other spatial variables such as accessibility, residential density, locational characteristics within the study area and types and mixture of businesses, which will potentially provide more complete information on the association between places and multiway social interaction (e.g., Yoon et al., 2011). Moreover, spatio-temporal distributions of activity participation should incorporate ideas such as the complexity of choice and satiation (Bhat, 2008; Sener et al., 2008) and measurable place attributes and attitudes such as sense of place (Deutsch et al., 2013a,b).

With respect to the kriging interpolation there is a variety of other experiments one can attempt to test spatial anisotropy and account for directionality in business distribution such as concentration along corridors, consider spatio-temporal covariance explicitly instead of converting time into distance, use of local kriging to account for the difference in distribution by locality, and perform comparisons with other spatial analysis methods as was done in the references cited in this paper in the kriging section.

Acknowledgements

Funding for this research was provided by The University of California Transportation Center, and the University of California Office of the President for the Multi-campus Research Program Initiative on Sustainable Transportation and the UC
Lab Fees program on Next Generation Agent-based Simulation. The authors would like to thank Michael Hindrichs and Kathleen Deutsch for rectifying coordinates. Anonymous reviewers provided extremely helpful comments and sources in the literature that helped provide a more complete review of contribution in this research field. Phaedon Kyriakidis and Chris Funk of the Department of Geography at UCSB are also thanked for their valuable inputs in Geostatistics. This paper does not constitute a policy or regulation of any Local, State, or Federal agency.

References


