Analytic Methods in Accident Research 10 (2016) 26–46

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

journal homepage: www.elsevier.com/locate/amar

A spatially autoregressive and heteroskedastic space-time pedestrian exposure modeling framework with spatial lags and endogenous network topologies

Jungyeol Hong, Venky N. Shankar*, Narayan Venkataraman

Department of Civil and Environmental Engineering, Pennsylvania State University, University Park 16802, USA

ARTICLE INFO

Article history:
Received 5 February 2016
Received in revised form
4 May 2016
Accepted 5 May 2016
Available online 14 May 2016

Keywords:
Pedestrian exposure
Network topology
Spatial autoregressive
GMM estimation

ABSTRACT

The main objective of this study is to derive a modeling framework for characterizing the space-time exposure of pedestrians in crosswalks, where the spatial measure is characterized by pedestrian density and the temporal measure is characterized by crosswalk time occupancy. This characterization has not been observed in the literature, but is a characterization that allows one to differentiate the components of pedestrian exposure with enhanced resolution in space and time. However, real-time observations to generate space-time data are time consuming and expensive over a large urban network. A hybrid microsimulation-statistical approach is utilized for data generation and statistical analysis in this study. The exposure models predicting crosswalk density and occupancy were estimated using spatial autoregressive models with spatial lags, autoregressive and heteroskedastic spatial disturbances and endogenous regressors. An instrumental variables generalized method of moments (IV-GMM) approach was used for estimation, and the spatial models account for spatial dependence among crosswalks through the estimation of spatial lag and spatial correlation parameters. In a case study of the downtown crosswalk grid in Seattle, Washington, 688 crosswalks were modeled using ten network topology measures capturing node degree, centrality, clustering, modularity, attractiveness and eccentricity measures. The models utilized these network topology variables to account for stochasticity in network design effects on pedestrian dynamics. Several important findings resulted from this study. First, and most important, it was found that network topology measures had an endogenous impact on pedestrian density. Second, the pedestrian time occupancy equation is characterized by endogenous selection effects. That is, in crosswalks with persistent pedestrian volumes and positive densities, the impact of pedestrian trip generation volumes and pedestrian density were corrected for endogeneity and selection bias. The combined results of the pedestrian density and time occupancy equations indicate that endogeneity and selection bias are critical issues that should not be ignored in pedestrian exposure modeling. Pedestrian trip generation volumes representing block level facility generation were found to be elastic. This finding indicates the utility of our modeling framework for estimating the impact of land use on pedestrian space-time exposure at the block level. Out-of-sample prediction tests of the density and time occupancy models and comparisons with pedestrian count data from field observations indicated substantial predictive accuracies. Finally, it was determined that degree and hub were highly sensitive network design parameters in terms of their influence on density. The average total impact (marginal effect) of these measures indicates that attention should be paid to crosswalk network design from the standpoint of

* Corresponding author.
E-mail addresses: juh255@psu.edu (J. Hong), shankarv@engr.psu.edu (V.N. Shankar), nsv111@psu.edu (N. Venkataraman).

http://dx.doi.org/10.1016/j.amar.2016.05.001

degree and hub characteristics. These results show that our space-time density-occupancy modeling framework is a plausible and efficient predictive tool that can be used to estimate pedestrian crosswalk exposure using *building level and network topology data* alone. We find that the IV-GMM technique is a useful approach for the emergent problem of inference in hybrid simulation-statistical transportation datasets, due to fewer assumptions on distributional assumptions about the data, while accounting for statistical effects relating to endogeneity, potential selection effects and heteroscedasticity.

1. Introduction

Emerging work in the development of sustainable transportation systems has led to an increased emphasis on the importance of non-motorized transportation. Importantly, pedestrian exposure is emerging as a significant measurement issue in the estimation of pedestrian-vehicle safety models. Pedestrian volume at the hourly or higher levels has been used historically as a key measure for exposure, but the spatio-temporal dynamics of pedestrian exposure remains largely unaddressed. Several studies have proposed the number of pedestrians or pedestrian volume to be a significant factor for pedestrian exposure estimation (Cameron and Milne, 1978; Tobey et al., 1983; Raford and Ragland, 2006). Conversely, walking distance has also been considered as a measure of exposure (Kaplan, 1975), and a similar study by Silcock et al. (1998) defined pedestrian exposure through vehicle travel distance (km), with the pedestrian accident rate per exposure unit being set as a risk measure. Keall (1995) evaluated the risk to pedestrians with walking travel time and the number of crossings as measures of exposure, and the study by Chu (2003) focused on a time-based exposure measure which was estimated using trip durations. Recent work on spatio-temporal patterns of pedestrian travel activity has focused on the harvesting of cell phone data (Bahoken and Raimond, 2013), and other work conducted recently with respect to downtown urban grids (Kwon and Shankar, 2014; Shankar and Kwon, 2013) lays out a methodology that can provide coarse-grained spatial-temporal security risks. However, the published literature has not considered the explicit role of network topology on pedestrian exposure from a safety standpoint. In particular, network topology effects relating to connectivity of crosswalks, spatial separation and the localized importance of certain crosswalks in an urban grid remains unaddressed. Ignoring network topology effects on pedestrian exposure can lead to propagation of error in pedestrian safety estimation. Since, pedestrian exposure is used as a volume-based variable in pedestrian crash models, this variable when estimated (in the absence of observed data) needs to be unbiased and accurately predicted. To this end, the estimators in the exposure model should not suffer from endogeneity, effects of spatial correlation and heteroscedasticity.

The objective of this study is therefore to develop pedestrian exposure models with spatial-temporal considerations that can provide for an understanding of network topology effects while enabling prediction at multiple scales (from 5 min intervals to hourly level), while accounting for endogenous network topology effects, heteroscedasticity, and spatial correlation. The rest of this paper is organized as follows: we begin with a description of the spatial scope of this study and the data bases used for the development of the models; followed by a description of the analytical process for model development; and a description of network topology and pedestrian flow measures used in the study. We then discuss the model development methodology, followed by a discussion of model findings and conclusions and directions for further research.

2. Data development

2.1. Spatial scope of the study and network definition

The spatial scope of this study consists of 50 traffic analysis zones (TAZs) in the downtown region of Seattle in Washington State. The 50 TAZs have a total of 3622 *facilities* that were utilized in trip generation calculations towards building an origin-destination matrix for pedestrian flows. Facilities were comprised of *building structures, sidewalks between two crosswalk endpoints, and crosswalks*. These components are areal units of analysis in our microsimulation. Table 1 shows a summary of the areal units for each corridor. For this study, since the exposure issue relates to crosswalks, we focus on model development for the crosswalk areal units in the network.

The trip exchanges between the 50 TAZs in this study and the remaining traffic analysis zones in the Greater Seattle area (1200 zones in total) were aggregated at an hourly level or higher (for example, 6 AM to 9 AM; 9 AM to 3 PM, and 3 PM to 6 PM). The study area was divided into 11 corridors representing consistency of land use (recreational versus commercial office versus institutional, etc.) while ensuring block-size homogeneity. (For a detailed description of the study area, see Kwon and Shankar (2014)).