Network Based Point Pattern Analysis of Traffic Accidents in City of Cape Town, South Africa

Christo du Toit · Sulaiman Salau · Sebnem Er

University of Cape Town, South Africa
Department of Statistical Sciences

17th Applied Statistics International Conference

20-22 September 2021

Introduction

Spatial point pattern events along networks can be observed in many different areas of human activity. In this research paper we focus on the traffic accidents that occurred in the City of Cape Town metropolitan municipality in 2016, and we estimate the intensity of accidents using both planar and network based kernel density (KDE) functions.

Objectives

The main aim of this project is to estimate the density of road traffic accidents in Cape Town along a network using the network KDE method. In section 2, a literature review of the studies with the same aim are provided. Section 3 presents the study area and the dataset where Section 4 provides details of planar and network density estimation methods. Results from our analyses are summarised in Section 5 and finally Section 6 concludes the findings and future research.

Literature

There are numerous studies estimating the density of spatial events that occur on a network such as crimes, economic activities and road traffic accidents in major city road networks etc. Most of these studies usually perform both planar and network KDE and make comparisons between the two commonly used methods [1, 2, 3].

References

[1] Zhixiao Xie and Jun Yan. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of Transport Geography*, 34:64–71, 2013.

[2] Timothée Produit, Nicolas Lachance-Bernard, Emanuele Strano, Sergio Porta, and Stéphane Joost. A network based kernel density estimator applied to barcelona economic activities. In David Taniar, Osvaldo Gervasi, Beniamino Murgante, Eric Pardede, and Bernady O. Apduhan, editors, *Computational Science and Its Applications - ICCSA 2010, International Conference, Fukuoka, Japan, March 23-26, 2010, Proceedings, Part I*, volume 6016 of *Lecture Notes in Computer Science*, pages 32–45. Springer, 2010.

[3] Yue Tang, Michael A.Knodler, and Mi-Hyun Park. A comparative study of the application of the standard kernel density estimation and network kernel density estimation in crash hotspot identification. In *16th Road Safety on Four Continents Conference*, number 16, 2013.

[4] Noel Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.

[5] Michal Bíl, Richard Andrášik, and Zbyněk Janoška. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident Analysis & Prevention*, 55:265–273, 2013.

[6] StatsSA. Statistics by place 2011. Available at http://www.statssa.gov.za/?page_id=993id=city-of-cape-town-municipality (2021/08/27), 2011.

[7] Frank H. Stillinger and David R. Herrick. Bound states in the continuum. *Phys. Rev. A*, 11:446–454, Feb 1975.

[8] J. von Neumann and E. P. Wigner. *Über merkwürdige diskrete Eigenwerte*, pages 291–293. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993.

Planar & Network Based KDE

One of the most commonly used method for estimating the density of road traffic accidents is Kernel density estimation which is usually performed in two forms: (1) Planar KDE and (2) Network KDE.

Earlier studies used planar Kernel density estimation (KDE) to estimate the density of traffic accidents. Planar KDE is based on the Euclidean space where spatial points events are weighted according to their radial distances (bandwidth) from the grid centroid. Events inside a specified radial bandwidth contribute to density estimation.

Let $(\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n)$ be the spatial locations of n accidents in the City of Cape Town bounded region (A) and $\mathbf{s}$ is any location on the network. The intensity $\lambda(.)$ of the accidents is estimated using the $\kappa(.)$ probability density function as follows [4]:

$$\hat{\lambda}_\tau(\mathbf{s}) \equiv \sum_{i=1}^{n} \frac{1}{\tau^2} \kappa \left(\frac{\mathbf{s} - \mathbf{s}_i}{\tau} \right), \qquad \mathbf{s} \in A, \qquad (1)$$

where $\kappa(.)$ is the kernel weighting function therefore the intensity estimation depends on the choice of the kernel and the bandwidth (τ). There are numerous choices for kernel weighting function however the most commonly used on is the quartic kernel function.

However, since road traffic accidents occur along a road network, which is composed of lines and vertices, the analysis of the density of traffic accidents and identification of hotspots along the network with planar KDE was found not to be sufficient. More recently, network KDE which is based on the network constrained space where events are weighted according to the distance measured along the network has been developed to improve the density estimation. In network KDE, projected events along a network bandwidth contribute to density estimation and it has been shown that it provides more accurate density estimation than planar KDE.

In both approaches, appropriate bandwidth determination is the key for the density estimation and [5] recommends a bandwidth between 100 and 300 meters for urban areas.

Study Area and Data

We focus on the traffic accidents that occurred in the City of Cape Town metropolitan municipality in Cape Town, South Africa. Cape Town is the second largest city in South Africa with 28.6% of the population within the municipality according to the national census 2001 and the second most important contributor to national employment [6].

The study area consists of wards in the city-bowl of the Cape Town Municipality area (Bokaap, Tamboerskloof and Greenpoint) with well-developed road network (Figure 1). Data on road traffic accidents was sourced from the City of Cape Town. The full data set contains records of more than 200 000 road traffic accidents that occurred between 2015-2017. The accident addresses were geocoded to obtain latitude and longitude coordinates which were then snapped to a connected road network withing the study area.

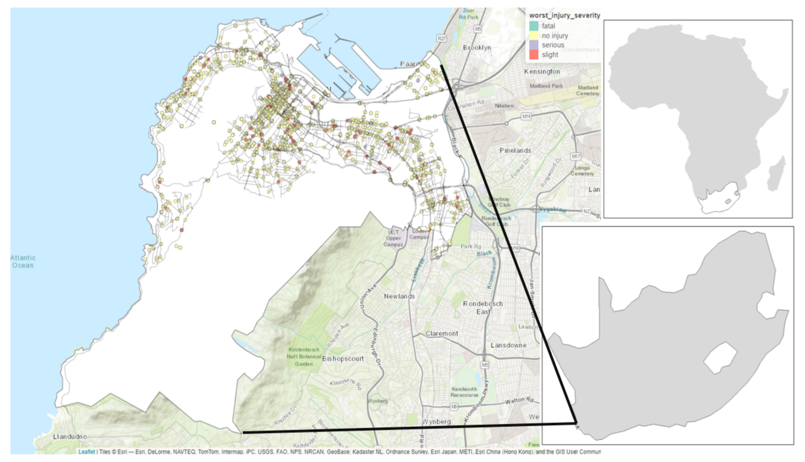

Figure 1: Study Area: City of Cape Town, South Africa - Traffic Accidents 2016

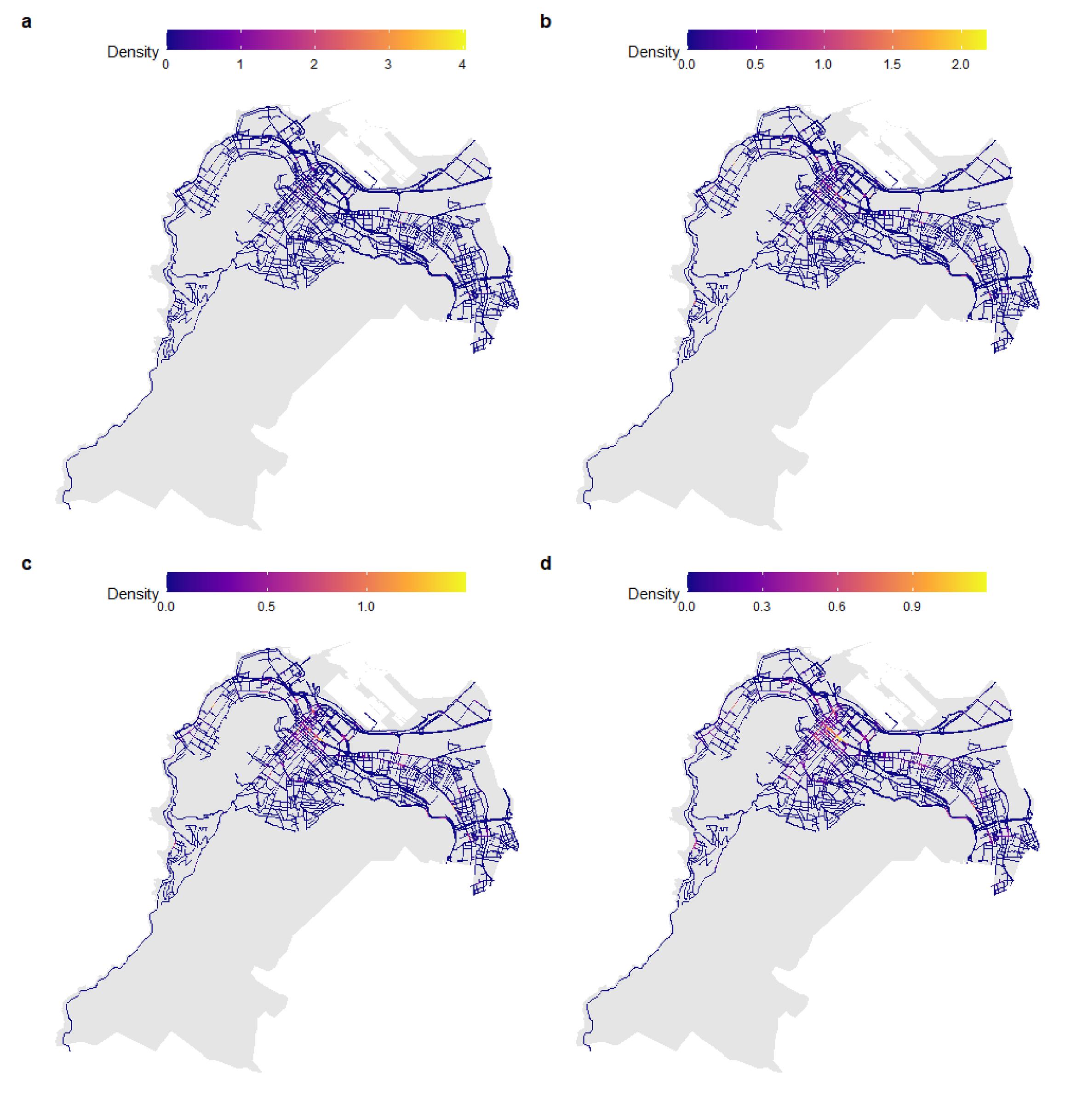

Figure 2: Spatial pattern of traffic accidents: density values calculated using a network KDE with varying bandwidths. a) 50m b) 100m c) 150m and d) 200m

Results and Discussion

Figures 2a-d show the density values calculated using a network KDE computed with various search bandwidth (50, 100, 150, and 250m). All analyses were conducted in R version 4.1.0 (R Core Team, 2021).

- Results show that Network KDE can better depict the nature and extent of accidents along the road network
- Network KDE influenced by the choice of bandwidths
- Another important input is the connectedness of the road network

Conclusion and Future Work

This study is one of the few attempts to comprehensively geocode accident data and explore the spatial distribution of accidents in the City of Cape Town. Next steps include:

- Improve geocoding match rates
- Expand study area and road types (connected and unconnected networks)
- Integrate exposure data for comparison of hotspots
- Fully access the impacts of kernel functions
- We can then start to suggest priority areas for improved safety of pedestrian and motorists