Spatial Data Analysis Using Kernel Density Tools

Boris A. Portnov
University of Haifa, Haifa, Israel

Marina Zusman
University of Haifa, Haifa, Israel

INTRODUCTION

Density analysis uses empirically recorded observations of some phenomenon or event (such as home locations of business customers, crime spots or locations of past forest fires) and transforms these observations into continuous surfaces of the "event density." Such surfaces show where the events of interest are concentrated geographically, thus helping to analyze their underlying patterns (Silverman, 1986; ESRI, 2012).

The main advantage of density analysis, compared to other popular methods of spatial smoothing (such as, spline, inverse distance weighted (IDW) method or *kriging*) is that it can be used for a study of a spatial phenomenon that has no recorded attributes, except for geographic location (Portnov, Dubnov, & Barchana, 2009).

In this chapter, we shall discuss the basic features of the density analysis method, focusing, in particular, on the Double Kernel Density (DKD) approach, which can be effective, as we shall demonstrate, for both mapping and multivariate analysis of data in a wide range of empirical applications.

THE USE OF DENSITY ANALYSIS IN EMPIRICAL STUDIES

In many business and research applications, there is a need to investigate the spatial distribution of some events of interests, in order to identify their geographic "hotspots" and to develop policy remedies, if required (see Table 1).

Thus, for example, in their study of the association between air pollution and lung cancer in the City of Haifa, Israel, Portnov et al. (2009), used the density method for a multivariate analysis of environmental and socio-demographic factors affecting lung cancer risk. In the studies two analytical methods – calculation of age-standardized rates (ASRs) for small census areas and density analysis – were compared. While AS comparison across small census areas detected no significant association between air pollution and lung cancer risk, density analysis helped to identify a significant association between lung cancer incidence and SO_2 pollution. The authors of the study demonstrated that density tools can be especially useful for the analysis of data that have only X,Y coordinates, such as, e.g., residential locations of lung cancer patients.

In another recent study by Gonzalez-Olabarria, Brotons, Gritten, Tudela, and Teres (2012), density analysis was used for the identification of regions in Spain with a recurrent history of forest fires. With the help of KD tools, the researchers compared the causes of ignitions and occurrence of hotspots. The analysis demonstrated that ignitions in the region of Catalonia were not random, thus helping to concentrate future firefighting efforts on specific "hotspots" of an elevated ignition risk.

Density analysis has been also widely used in criminology. Thus, Porter, and Reich (2012) used KD estimations to forecast the locations of future criminal and terrorist events. The research-

DOI: 10.4018/978-1-4666-5202-6.ch203

Table 1. Examples of density analysis use in different scientific fields and applications

Scientific Field	Research Application
Criminology	Forecasting locations of future criminal and terrorist events (Porter, & Reich, 2012)
	Mapping the addresses of recorded crime events for the identification of crime spots (Wolff, & Asche, 2009)
Environmental sciences	Estimating the conditional probabilities of the rainfall (Sharma, 2000)
	Identification of regions with recurrent forest fires (Gonzalez-Olabarria et al., 2012)
	Forecasting wind power potential (Taylor, & Jeon, 2012)
Epidemiology and public health studies	Identifying the association between disease density and environmental risk factors (Kloog, Haim, & Portnov, 2009; Portnov et al., 2009; Zusman, Dubnov, Barchana, & Portnov, 2012)
	Analysis of the distribution of alcohol outlets in residential neighborhoods (Carlos, Shi, Sargent,Tanski, & Berke, 2010)
Marketing	Analysis of the service area of a restaurant (Donthu, 1991)
	Modeling the patterns of customer density aimed at locating prospective new customers (Sliwinski, 2002)
Transportation	Density analysis of recorded traffic accidents for understanding their spatial patterns (Anderson, 2009; Xie, & Yan, 2008)
	Space–time analysis of traffic trajectories of passenger ships and tankers (Demšar, & Virrantaus, 2010).

ers based their analysis on a temporally weighted KD model, used to predict how much influence past events may have on predicting future event locations. The analysis showed that forecasting quality tends to vary by the type of crime and relatively little by time-series length.

TYPES OF DENSITY ESTIMATES

There are two basic approaches to calculating density surfaces – *linear averaging* and *kernel density (KD)*. In both methods, the areal density of geographically referenced features (or events) is calculated for a neighborhood of a raster cell (pixel), which "net" covers the entire area of interest. By way of this estimation, density is calculated as the number of features or events located in each cell's predefined proximity range, per unit of area (that is, per m^2 or km^2 or any other areal measure). To this end, a circle is drawn around each raster cell, using some predefined search radius (see the subsection on the "KD calculation," for more detail). The number of features (or events), which fall inside the circle, are then summed up, and the total number of events is divided by the circle's area (Gatrell, Bailey, Diggle, & Rowlingson, 1996; Silverman, 1986). KD estimates are calculated similarly, but according to this method, points, lying near the center of a given raster cell, are weighted more heavily than those lying near its edge (McCoy & Johnston, 2001).

Figure 1 helps to illustrate differences between maps produced by two different density methods - linear averaging (Figure1 A) and KD method (Figure1 B). In both cases, the same geographic distribution of input points (events) is used, with these points being marked by tiny black dots on the maps. In particular, as Figure 1 shows, the calculated density surface, estimated using the linear averaging (Figure 1A), appears to be coarser and more dispersed then the surface obtained using KD estimates (Figure 1B).

KD CALCULATION

The KD calculation procedure is relatively simple; it is based on a non-linear kernel function, λ_τ estimated as follows: