

Hotspot mapping using Kernel Density Estimation (KDE)

Powerful visualisation and analysis of incident data

A Cadcorp application for mapping crime data and other socio-economic data sets as smooth continuous surfaces

Hotspot mapping has become a valuable technique for visualising the geographic incidence of crime and other socio-economic data. One of the most widely used techniques for generating hotspot maps as smooth continuous surfaces is kernel density estimation (KDE). Cadcorp now offers this technique as standard in the Cadcorp SIS family of desktop GIS products. This document describes the capabilities of the Cadcorp implementation of KDE.

What is hotspot mapping?

Unlike point mapping which focusses on mapping the location of individual events, hotspot mapping focusses on highlighting areas which have higher than average incidence of events. These have come to be known as 'hotspots'. The technique has been shown to have advantages over other methods of aggregating point data, such as spatial clustering, choropleth mapping, and quadrat mapping.¹ One of the key differentiators is that hotspot mapping represents the incidence of events as a smooth continuous surface which, generally accords with most users' expectations.

What is kernel density estimation (KDE)?

KDE provides an estimate of the proportion of total incidents that can be expected to occur in any given map location. It works by first overlaying an area of interest with a fine rectangular grid. It then calculates an estimate of the density of incidents in each

grid cell which is based on a weight function - the kernel. The kernel is a function of specified shape and bandwidth (or search radius). It is defined to reflect the belief that there is a greater probability of an incident occurring in a given location the closer it is to the location of a known incident.

We can visualise the kernel function as describing a 'mound-shaped' volume which we centre at the location of each incident in our data set. The influence of the incident on any grid cell location is measured by the height of the mound's surface at the centre of that grid cell. Where incidents are clustered, their respective mounds may overlap. The final density surface is calculated by taking account of the aggregate influence of all incidents by summing the heights of the mound surfaces at each cell.

What parameters can be specified?

Cadcorp's implementation of KDE allows for three parameters to be specified. These are the following:

- grid cell size
- bandwidth (the search radius of the kernel density function), and
- the shape of the kernel density function

Other things being equal, a finer grid cell size will provide more visually appealing smooth surfaces. However, this may be at the expense of data processing speeds, as larger data volumes are involved. A large bandwidth will spread the influence of

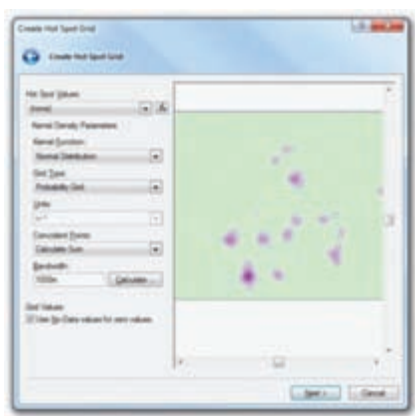
¹See for example 'Mapping Crime: Understanding Hot Spots', U.S. Department of Justice, National Institute of Justice Special Report, Aug '05

points over a greater area. This may be more appropriate for strategic analysis, but less appropriate for local analysis. Too large a bandwidth will suppress spatial variation. If it is too small, it will not produce a continuous smooth surface. A flatter kernel function will also result in a more uniform distribution. In general the bandwidth of the kernel function density estimation often proves to be more influential to outcomes than the kernel shape.

Wizard-based interaction

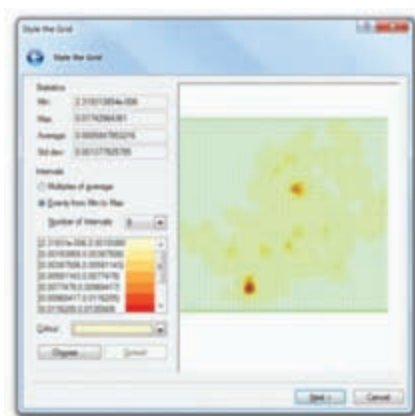
A wizard guides the user through the several stages involved in generating a hotspot map. The key stages are discussed below.

The wizard begins by guiding the user through the specification of grid cell size and origin together with the shape and the bandwidth of the kernel function to be applied. The user can select from 9 different kernel functions.



Creating the hot spot grid

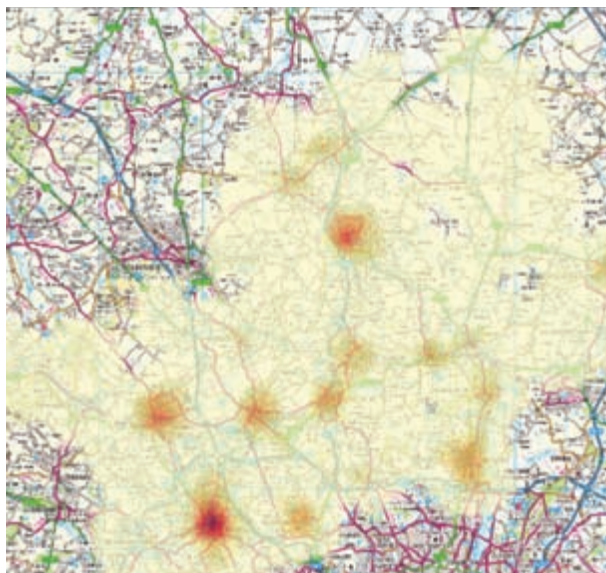
Cadcorp SIS also allows the user to specify the type of grid: probability, relative density, and absolute density. An option exists to specify how coincident points are treated (as sum, average, minimum, or maximum). The determination of an appropriate bandwidth value is helped by the provision of an option to calculate K-mean Nearest Neighbour distance.



Styling the grid

After defining the settings for the Hotspot Grid, the wizard guides the user through a thematic styling dialogue. The user can specify the number and extent of map intervals, based on inspection of the statistical distribution of grid values. Alternatively, the application can be allowed to set values automatically.

Grids can be stored as part of an existing map overlay, or alternatively, held as a Grid Overlay in its own right. The user can apply colour sets to the resulting analysis from a variety of pre-defined palettes.



A 2-D view of a hot-spot map overlaid on a base map.



Hot spot maps can also be viewed in 3-D

The strengths of Cadcorp Hotspot Mapping using KDE

The Cadcorp implementation of KDE is distinctive for the following reasons:

- it supports easy hot-spot mapping of any socio-economic event data
- it allows for three parameter settings: grid cell size, bandwidth, and kernel density function
- a wizard-based workflow allows the user to refine output with ease
- the KDE analysis capability is included as an integrated component of the Cadcorp desktop product family at no extra cost