[AM-03-008] Kernels and Density Estimation

Abstract

Kernel density estimation is an important nonparametric technique to estimate density from point-based or line-based data. It has been widely used for various purposes, such as point or line data smoothing, risk mapping, and hot spot detection. It applies a kernel function on each observation (point or line) and spreads the observation over the kernel window. The kernel density estimate at a location will be the sum of the fractions of all observations at that location. In a GIS environment, kernel density estimation usually results in a density surface where each cell is rendered based on the kernel density estimated at the cell center. The result of kernel density estimation could vary substantially depending on the choice of kernel function or kernel bandwidth, with the latter having a greater impact. When applying a fixed kernel bandwidth over all of the observations, undersmoothing of density may occur in areas with only sparse observation while oversmoothing may be found in other areas. To solve this issue, adaptive or variable bandwidth approaches have been suggested.

Keywords: bandwidth, basic analytical methods, data smoothing, density map, hot spot detection, kernel, line data, local density, point data

Author & citation

Yin, P. (2020). Kernels and Density Estimation. The Geographic Information Science & Technology Body of Knowledge (1st Quarter 2020 Edition), John P. Wilson (ed.). DOI: [10.22224/gistbok/2020.1.12](https://doi.org/10.22224/gistbok/2020.1.12).

An earlier version can also be found at:

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). Kernels and density estimation. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers.

Explanation

1. Definitions

Density: The amount of features or events per unit area

Density estimation: The construction of the density function from the observed data

Kernel: A window function fitted on each observation (weighted or unweighted) to determine the fraction of the observation used for density estimation at any location within the window

2. Why Do We Need Kernel Density?

Density calculates the amount of features or events per unit area. It is an important measure of the distribution pattern of features or events. For example, population density can be used to identify where people cluster in a region. Zone-based density is the simplest and most widely- used density measure in the context of GIS&T, which applies some form of zoning, such as a regular grid or a census, to a region. For each zone, the density is then calculated by dividing the total amount of the features or events within the zone by the size of the associated zone.

Zone-based density is easy to understand and to calculate. However, it has several limitations. First, the delineation of the zones determines the zonal size and how the features or events are aggregated over the zones. Therefore, for the same set of observed data, the densities could vary substantially depending on the delineation of the zones used, which leads to a difficulty in the interpretation of the distribution pattern. This is a manifestation of the well-known modifiable areal unit problem (Openshaw & Taylor, 1981) in spatial analysis. Secondly, each zone has a constant density and the local variation within the zone is obscured. Thirdly, a zone-based density surface generally is not continuous as density may change abruptly at the zone boundaries.

As an alternative to zone-based density, kernel density can solve or mitigate the above issues by constructing a continuous density surface from the observed data. Kernel density estimation has been widely used for various purposes, such as point or line data smoothing, risk mapping, and hot spot detection. In public health and health geography, kernel density has been used to measure spatial accessibility to health services (Yang, Goerge, & Mullner, 2006), estimate disease incidence rates and identify hot spot areas for intervention targeting (Rushton et al. 2004), and investigate the relationships between environmental exposures and health outcomes (Kloog, Haim, & Portnov, 2009; Portnov, Dubnov, & Barchana, 2009). In archeology, kernel density has been used to derive historical land-use surfaces based on archeological sites (Bonnier, Finné, & Weiberg, 2019), to estimate the densities of geomorphological and geological hazards for understanding the positioning constraints of historical buildings (Mariani et al., 2019), and as a classification tool to group specimens based on the analytical results of the chemical composition (Santos et al., 2006).

3. Point Kernel Density Estimation

Kernel density can be estimated from point-based or line-based data. Given its simplicity and popularity, this section only focuses on kernel density for point-based data where points are assumed to be distributed across a continuous space. Some basic information about kernel density for lines and networks can be found in Section 4, below.

Kernel density estimation fits a kernel, i.e., a window function, on each observation point (weighted or unweighted). Similar to zone-based density that assumes observation points are continuously distributed within their surrounding zones, kernel density estimation assumes each observation point is continuously spread within its fitted kernel window.

However, different from zone-based density that simply assumes an even distribution of observation points within the zones, kernel density estimation calculates the fraction of an observation point at location **x** based on the kernel function used and the distance between location **x** and the observation point. For each location in the space, the kernel density (i.e., observations per unit area) is calculated by summing up the fractions of all observation points at that location. If dividing the kernel density by the total number of observation points, the result would be the probability density estimate indicating the chance that an observation could occur at that location. Eq.1 defines the kernel density estimate $\hat{f}$ at location **x** from a set of weighted observation points **X1 ,,, Xn** . Each **wi** represents the weight of **Xi**, such as the number of crimes or the total population at that location. **K** is the kernel function that determines the fraction of the weight of **Xi** at location **x** based on the distance **di** between the two points (i.e., **x - Xi**). Eq.1 can be also applied to unweighted observation points by setting **w1 , ... , wn** with the value of one.

$$\hat{f}(x) = \sum_{i=1}^{n} w_i K(d_i) = \sum_{i=1}^{n} w_i K(x - X_i) \qquad (1)$$

For simplicity, Fig.1 illustrates an example of one-dimensional (univariate) kernel density estimation from five unweighted observation points. In this example, a standard normal (Gaussian) kernel (black dashed curve) is fitted over each observation point so that the fraction of the point is the greatest at the point location and diminishes when moving away from the point. The red solid curve shows the estimate of the density function whose integral equals to the total number of observation points that is five in this case, and the blue solid curve shows the estimate of the probability density function whose integral equals to one.

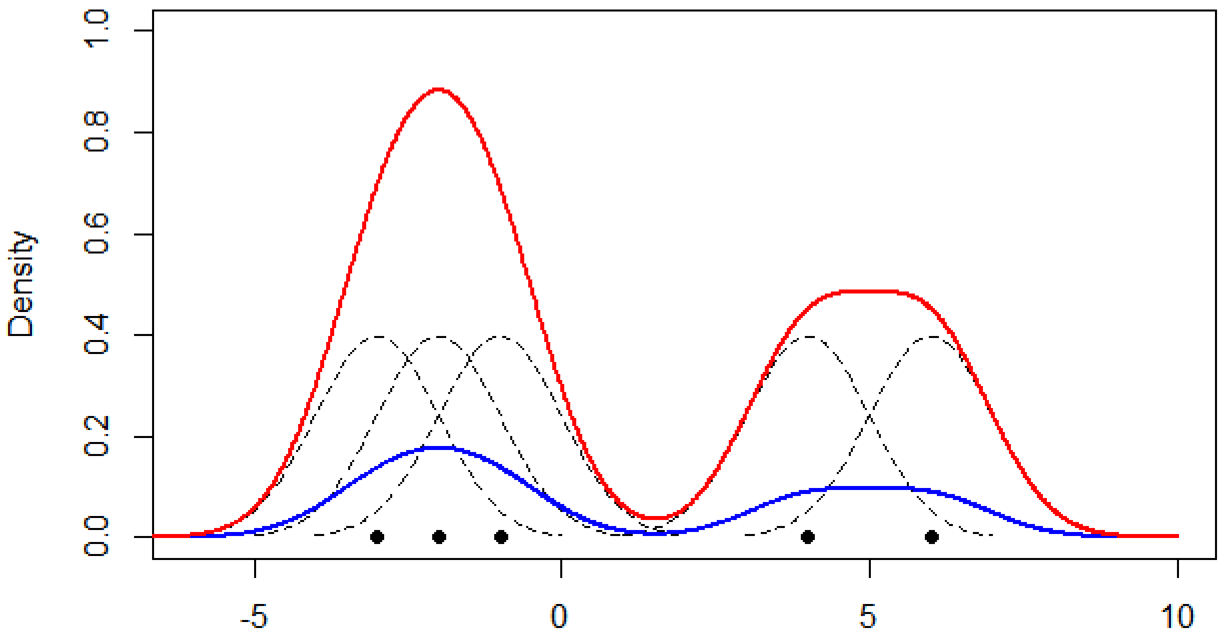
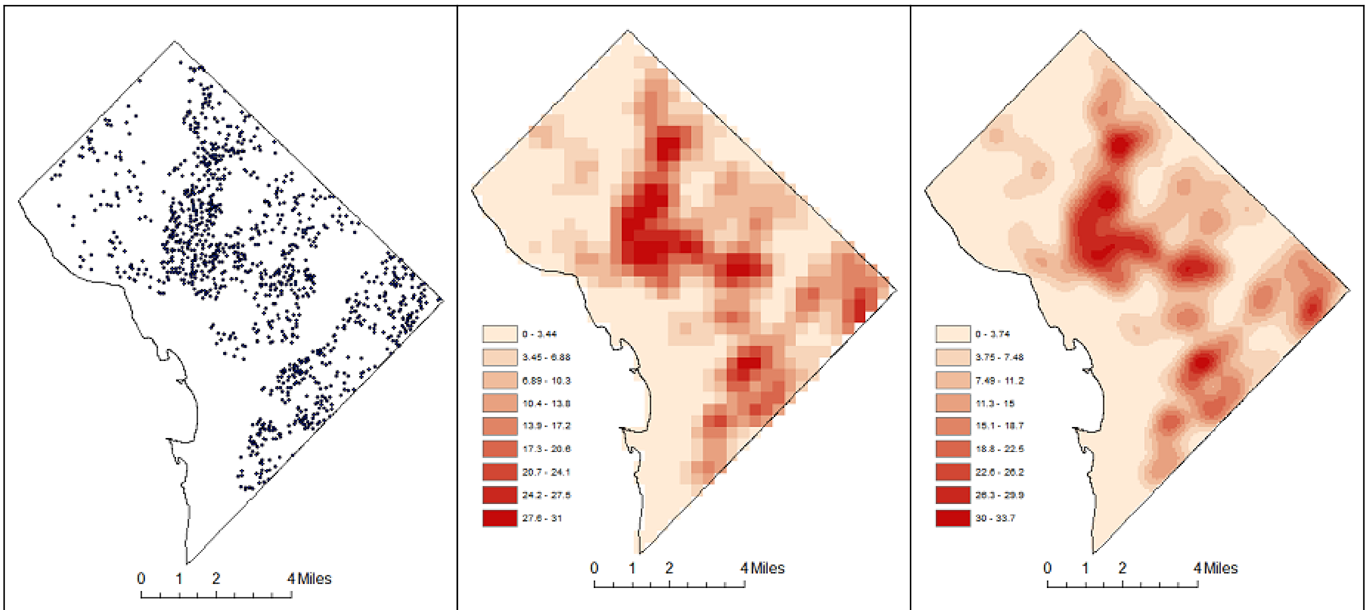


Figure 1. Example of one-dimensional (univariate) kernel density estimation. Source: author.

Spatial analysis usually deals with two-dimensional space. For two-dimensional (x, y) point data (weighted or unweighted), such as crimes or disease cases plotted on a planar map, the kernel density estimation procedure is almost the same as univariate density estimation except that a second dimension needs to be added to the kernel. When a symmetric kernel is used, the procedure essentially can be thought as rotating the kernel on each point around its central axis and then scaling the resulting kernel window to make the total volume under the kernel surface equal to one.

Theoretically, kernel density can be estimated at any location. However, it is impossible for a computer to calculate the density estimate for every point in the space because the number of points is infinite. A common way is to overlay a grid on top of the space and calculate the density estimate for the center point of each grid cell. A smaller cell size usually leads to a more detailed density surface, but also makes the computation more intense. As an example, the kernel density estimate of burglary incidents in Washington DC in 2018 is calculated with two different output cell sizes: 100m and 500m. The burglary incidents data is downloaded from the website of Open Data DC and the DC boundary data comes from the US Census. The calculation was conducted with the kernel density tool in Esri’s ArcGIS for Desktop v10.6. All default values are taken for the parameters of the tool except for the output cell size. Fig.2(a) shows the original 1415 burglary incident locations. It should be noted that, for the purpose of confidentiality protection, these locations are the associated street block centers instead of the real crime locations. Fig.2 (b) and (c) show the kernel density estimation outcomes using two different output cell sizes. Obviously, the density surface with the output cell size of 100m (Fig.2 (c)) is much more smoothed and detailed than that with the output cell size of 500m (Fig.2 (b)). With the kernel density maps shown in Fig.2, we can see that the crimes were mainly concentrated in the center and southeast of Washington DC. This finding can assist the planning of crime prevention and control activities.



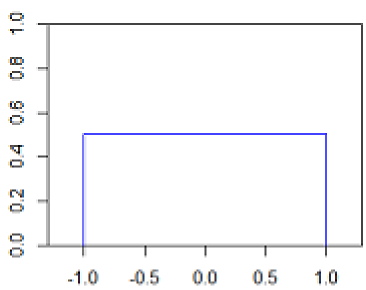
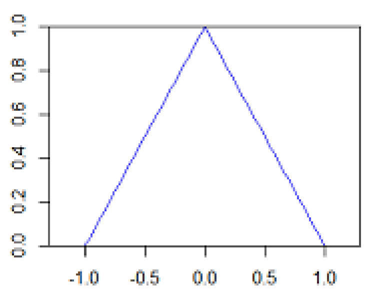
(a) burglary incident points | (b) kernel density estimate (output cell size = 500 m) | (c) kernel density estimate (output cell size = 100 m)

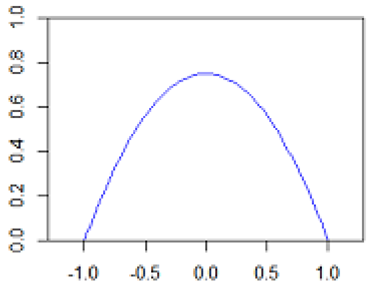
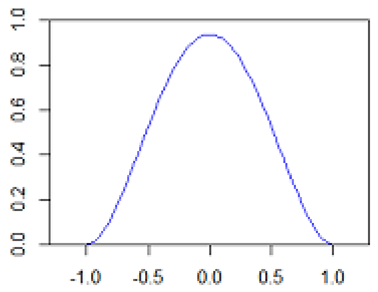
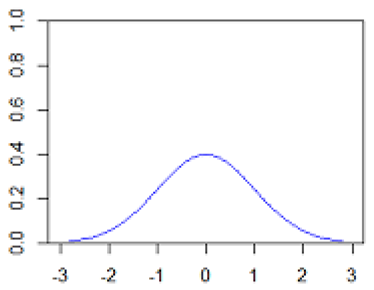
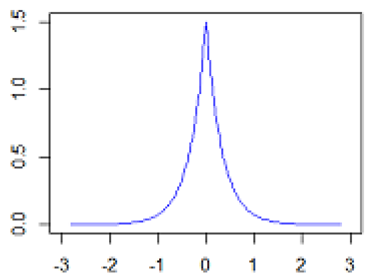
Figure 2a - 2c. Kernel density estimates of burglary incidents in Washington, D.C., 2018. Source: author.

3.1 Kernel Examples and Its Selection

A critical component in kernel density estimation is how an observation point is spread over the space, which is determined by kernel functions. A kernel function **K(d)** describes the distribution of an observation point with the distance **d** measured from the observation point. Usually, a kernel is a symmetric, non-negative function whose integral equals to one, such as a probability density function or a weight function meeting the integral constraint. Table 1 lists six types of standard kernel functions (the integral equals to one) that are commonly used in univariate density estimation. Each has a different shape determining how the observation point is spread. All of these standard kernel functions are bounded within one unit of the observation point except for the normal function and negative exponential function that are unbounded. The kernel density tool in Esri's ArcGIS for desktop v10.6 only supports quartic kernel for two-dimensional space (Esri, 2019), while the kernel density tool in the CrimeStat IV software supports five types of kernels for two-dimensional space, including normal, uniform, quartic, triangular, and negative exponential (Levine, 2013).

Table 1. Standard Kernel Functions Commonly Used in Univariate Density Estimations

Name	Formula	Illustration
Uniform (flat)		
Triangular (conic)		

Name	Formula	Illustration
Epanechnikov (paraboloid / quadratic)		
Quartic (biweight)		
Normal (Gaussian)	$K(d) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d^2}$	
Negative (exponential)	$K(d) = Ae^{t\|d\|}$ (where A is a constant and $t < 0$)	A = 3/2, t = -3 

Since most of the commonly-used standard kernels are bounded within one distance unit, to accommodate the data, they are usually scaled so that the observation point can be spread farther or closer. A scaled kernel **Kh(d)** can be defined as

$$K_h(d) = \frac{1}{h}K(\frac{d}{h})$$

(2)

where **K** is the standard kernel to scale and **h** is called bandwidth or smoothing parameter that controls the extent of the kernel scaling. It can be proved that the integral of **Kh** also equals to 1. The larger the bandwidth is, the farther the observation point spreads. More detail about bandwidth selection can be found in Section 3.2.

In the context of statistics, the observation points are assumed as a sample from an unknown probability density function. Density estimation is to construct an estimate of the density function from the observed data. It has been proved that, among all non-negative kernels with their own optimal bandwidth used, the Epanechnikov (paraboloid/quadratic) kernel is optimum in the sense of minimizing the discrepancy of the density estimator from the true density function. However, other commonly-used kernels, such as those listed in Table 1, can achieve very similar results (Silverman, 1986). Therefore, it is appropriate to base the choice of kernel on other considerations, such as the user's analytical experience, domain knowledge of the applications, and computational considerations. S. C. Smith and Bruce (2008) provides their reasoning on the choice of kernel for different types of crime. For example, they suggest the uniform kernel or other moderately dispersed kernels for residential burglaries because most burglars cruise neighborhoods looking for likely targets and a housebreak in any part of a neighborhood transfers risk to the rest of the neighborhood. Compared to residential burglaries, domestic violence incidents at one location are usually not contagious in the surrounding area, so a kernel that weights more heavily to the central point is suggested, such as negative exponential kernel.

3.2 Bandwidth Selection

Compared to the choice of kernel, the choice of bandwidth has a greater impact on the result of density estimation. Given a kernel, a smaller bandwidth tends to lead to an undersmoothed density estimate that may contain many spurious noises (i.e., peaks and valleys), and a larger bandwidth tends to result in an oversmoothed density estimate that may hide some interesting underlying structure. Therefore, choosing an optimal bandwidth for a dataset is important. However, there is no consensus on how to choose an optimal bandwidth (Levine, 2013). Theoretically, the optimal bandwidth of a kernel can be derived under the assumption that the optimal bandwidth can minimize the discrepancy of the density estimator from the true density. However, the derived optimal bandwidth using such a method depends on the true density that is usually unknown. Several practical methods have been developed to estimate the optimal bandwidth from a given dataset.

A natural method is to subjectively choose the bandwidth based on prior ideas about the true density. First, multiple density estimates with different bandwidths are plotted out. Chosen from these estimates, the optimal bandwidth is the one that leads to the estimate most in accordance with one's prior ideas about the true density. If the true density is totally unknown, which is the most common case, exploring multiple bandwidths may provide users some insight of the underlying data structure, which is useful for the purposes of model and hypothesis generation.

Another easy and natural method is rule-of-thumb bandwidth estimation. As mentioned above, the optimal bandwidth that minimizes the discrepancy of the density estimator from the true density depends on the true density. The rule-of-thumb method replaces the unknown true density by a reference distribution function that is re-scaled to have variance equal to the variance of the observations. Taking univariate kernel density estimation as an example, if the normal kernel is used for estimation and the normal distribution is used as the reference distribution for the true density, the rule-of-thumb bandwidth estimate **hrot** can be derived as

$$h_{rot} = 1.06\hat{\sigma}n^{-0.2} \quad (3)$$

where n and $\hat{\sigma}$ are the number and variance of the observations, respectively. The kernel density tool in Esri's ArcGIS (v10.2.1 or later) also uses the rule-of-thumb method with a quartic kernel to calculate the default bandwidth (search radius) for two-dimensional density estimation (Esri, 2019).

In addition to the subjective choice and rule-of-thumb method, there are several more sophisticated methods for optimal bandwidth estimation, such as cross-validation methods, plug-in methods, etc. More technical detail of these methods have been discussed by Silverman (1986) and Turlach (1993).

3.3 Fixed Kernel vs. Adaptive Kernel

When a fixed bandwidth is applied across the entire set of observations, it is possible that, in regions of low density, the kernel can only covers very few observations, while regions of high density will see a large amount of observations in the kernel. As a result, when a small fixed bandwidth is applied to data from long-tailed distributions, there is a tendency for spurious noise to appear in the tails of the estimates. If the estimates are smoothed sufficiently with a larger fixed bandwidth, then some essential detail in the main part of the distribution may be masked. To solve this issue, adaptive or variable bandwidth approaches, which vary the kernel bandwidth from one point to another, have been suggested.

Generally, there are two categories of adaptive kernel estimators. The first one is called balloon estimator that varies the kernel bandwidth depending on the test location. For example, to estimate the density at location x, the user can specify the bandwidth of a bounded kernel as the distance from location x to its kth nearest observation point. In other words, there are always k nearest observation points involved in the density estimation for each location. The second category of adaptive kernel estimators is called sample smoothing estimator. It varies the kernel bandwidth at each observation point. For example, the user can specify the bandwidth of the kernel at an observation point as the distance from the observation point to its kth nearest other observation point.

Terrell and Scott (1992) compared the two categories of adaptive kernel estimators with fixed kernel estimators in a simulation study. They found that adaptive estimators in a non-optimal fashion can be inferior to fixed methods for sufficiently large number of observations. Yet, the adaptive estimation could be particularly effective for multi-dimensional space, especially three or higher dimensions.

4. Kernel Density for Lines and Networks

Kernel density estimation can also be applied to linear features in a similar way as point kernel density estimation. Conceptually, a selected kernel is fitted over each line to spread the number, length, or other types of weight of the line over the space. Usually, the fraction of the line is the greatest on the line and diminishes as moving away from the line. The kernel density of linear features at a location is the sum of the fractions of all lines at that location. In addition to supporting point kernel density estimation, Esri's ArcGIS and R package "spatstat" both provide kernel density tools for line segments.

In the real world, many events happen on or alongside a network, such as car crashes on streets. Okabe, Satoh, and Sugihara (2009) have shown that the planar kernel density estimation methods discussed above are inappropriate to estimate the density of network points. To solve this issue, they proposed three types of kernel functions for network points and two of them are unbiased kernel density estimators when the true probability density is a uniform distribution. These network-based methods have been implemented in the SANET software.

References

Bonnier, A., Finné, M., & Weiberg, E. (2019). Examining land-use through GIS-Based kernel density estimation: a Re-Evaluation of legacy data from the berbati-limnes survey. Journal of Field Archaeology, 44(2), 70-83.

de Smith, M. J., Goodchild, M. F., & Longley, P. A. (2018). Density, Kernels and Occupancy. In: Geospatial Analysis: A Comprehensive Guide to Principles, Techniques, Software, and Tools (5th Edition).

Esri. (2019). ArcGIS Desktop v10.6 online help document: How kernel density works Retrieved Aug.15, 2019.

Kloog, I., Haim, A., & Portnov, B. A. (2009). Using kernel density function as an urban analysis tool: Investigating the association between nightlight exposure and the incidence of breast cancer in Haifa, Israel. Computers, Environment and Urban Systems, 33(1), 55-63.

Levine, N. (2013). Kernel Density Interpolation CrimeStat: A spatial statistics program for the analysis of crime incident locations (v4.0): N Levine & Associates, Houston, TX, and the National Inst. of Justice, Washington, DC.

Mariani, G. S., Brandolini, F., Pelfini, M., & Zerboni, A. (2019). Matilda's castles, northern Apennines: geological and geomorphological constrains. Journal of Maps, 15(2), 521-529.

Okabe, A., Satoh, T., & Sugihara, K. (2009). A kernel density estimation method for networks, its computational method and a GIS-based tool. International Journal of Geographical Information Science, 23(1), 7-32.

Openshaw, S. & Taylor, P. J. (1981). The modifiable areal unit problem. In N. Wrigley & R. Bennett (Eds.), Quantitative geography: A British view (pp. 60-69). London and Boston: Routledge and Kegan Paul.

Portnov, B. A., Dubnov, J., & Barchana, M. (2009). Studying the association between air pollution and lung cancer incidence in a large metropolitan area using a kernel density function. Socio-Economic Planning Sciences, 43(3), 141-150.

Rushton, G., Peleg, I., Banerjee, A., Smith, G., & West, M. (2004). Analyzing Geographic Patterns of Disease Incidence: Rates of Late-Stage Colorectal Cancer in Iowa. Journal of Medical Systems, 28(3), 223-236.

Santos, J., Munita, C., Valerio, M., Vergne, C., & Oliveira, P. (2006). Determination of trace elements in archaeological ceramics and application of Kernel Density Estimates: Implications for the definition of production locations. Journal of Radioanalytical and Nuclear Chemistry, 269(2), 441-445.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. New York: Chapman & Hall/CRC.

Smith, S. C., & Bruce, C. W. (2008). Chapter 6: Kernel Density Estimation. In: CrimeStat III: User Workbook. Washington, DC: The National Institute of Justice.

Terrell, G. R., & Scott, D. W. (1992). Variable Kernel Density Estimation. The Annals of Statistics, 20(3), 1236-1265.

Turlach, B. A. (1993). Bandwidth selection in kernel density estimation: A review. Paper presented at the CORE and Institut de Statistique.

Yang, D.-H., Goerge, R., & Mullner, R. (2006). Comparing GIS-Based Methods of Measuring Spatial Accessibility to Health Services. Journal of Medical Systems, 30(1), 23-32.