

A Comparative Study of Higher Order Kernel Estimation and Kernel Density Derivative Estimation of the Gaussian Kernel Estimator with Data Application.

Siloko Israel Uzuazor^{1*} and Ojobor Sunday Amaju².

* Corresponding Author



1. Department of Mathematics, Edo State University Uzairue, Nigeria. sisraelgreat@gmail.com, siloko.israel@edouniversity.edu.ng
2. Department of Mathematics, Delta State University, Abraka, Nigeria. ojoborsun@yahoo.com

Abstract

Higher-order kernel estimation and kernel density derivative estimation are techniques for reducing the asymptotic mean integrated squared error in nonparametric kernel density estimation. A reduction in the error criterion is an indication of better performance. The estimation of kernel function relies greatly on bandwidth and the identified reduction methods in the literature are bandwidths reliant for their implementation. This study examines the performance of higher order kernel estimation and kernel density derivatives estimation techniques with reference to the Gaussian kernel estimator owing to its wide applicability in real-life-settings. The explicit expressions for the bandwidth selectors of the two techniques in relation to the Gaussian kernel and the bandwidths were accurately obtained. Empirical results using two data sets obviously revealed that kernel density derivative estimation outperformed the higher order kernel estimation excellently well with the asymptotic mean integrated squared error as the criterion function.

Key Words: AMISE, Bandwidth, Density Derivative, Gaussian Estimator, Higher Order Kernel.

Mathematical Subject Classification: 62G05, 62G08, 62G20.

1. Introduction.

The Gaussian kernel estimator is of numerous applications in many fields of study. The extensive applicability of the Gaussian distribution is attributed to its ability to define the probability of any given data set and the smoothness of its estimates using mathematical tools. In probability theory, the Gaussian distribution explains one hundred percent of the values of observations to be investigated in a given probability space with the symmetric property. The Gaussian function is very important in image production owing to the fact that it can be easily extended to any desired dimension of interest. In density estimation, the Gaussian distribution is fundamental in probability estimation, especially in statistics and other related fields (Nan and Ji, 2020; Siloko et al., 2020a; Johnson et al., 2021).

Density estimation is the production of probability estimates from a set of observations and it is of great significance in statistical fields of studies. Generally, the probability density estimate may either be constructed from a known probability distribution which is regarded as parametric estimation or from unknown distribution which is nonparametric estimation. There is usually a predetermined structure in parametric estimation but structure pre-determination does not exist in nonparametric estimation because they are pliable. However; the pliability of nonparametric techniques is associated with a high computational cost which constricted their utilizations before the advent of fast computing systems. The cost of computational analysis of manifold observations is often connected with nonparametric estimation particularly with composite statistical designs. There is no imposition of distribution in nonparametric estimation because the data reveals the statistical composition which accounted for their extensive uses since information about some pre-historical data may not be readily available (Green et al., 2015; Scott, 2015).

One of the most studied nonparametric density estimation techniques is the kernel method and the prominence of the method is accredited to the clarity of its results and graphical presentation of observations in a comprehensible form for the purpose of visualization (Stanislaw, 2018). The kernel estimator is a probability function with several uses in many fields of study particular in this information age where series of big data are been analyzed for better decision making by individuals and organizations. The popularity of the kernel estimator is also attributed to its reckoning cost-effectiveness and clarity of presentation of data in a distinctive manner. As an exploratory and visualization tool, the kernel estimator has been used indirectly in many areas (see Raykar et al., 2015; King et al., 2016; Helu et al., 2017; Rojas, 2017; Spencer, 2017; Bonnier et al., 2019; Siloko et al., 2019; Li et al., 2019; Siloko et al., 2021). Despite the relevance of the kernel estimator in numerous fields and its wide applicability in density estimation, the complexity connected with the appropriate selection of the bandwidth or smoothing parameter which determines the level of smoothness of the estimate is a serious setback to its implementation. The smoothing parameter influences the performance of the kernel estimator but applying diverse kernel estimators like the Epanechnikov against Triweight kernel will definitely produce dissimilar outcomes due to variation in the derivatives of the function.

Nonetheless, the performance of kernel method in data analysis can be enhanced by utilizing higher order kernel and kernel density derivatives because of their bias and variance reduction property. Regarding higher order kernels, the performance typically rely on the magnitude of the bandwidth with its order due to the fact that they converge at faster rates than lower order kernels. The justification for higher order kernel is occasioned by their convergence rates but often constrained by the production of estimates that are with negative components hence, they are oftentimes not probability density estimate and this situation is always complex to be analyzed by data scientists. Irrespective of the negative estimates of higher order kernel, the estimation of massive data with higher order kernel has shown their superiority over the corresponding lower kernel. The improvement in performance of higher order kernel is regularly characterized by a large bandwidth contrary to the usual bandwidth in the estimation of lower order kernels (Marron and Wand, 1992; Jones, 1992; Marron, 1994; Wand and Jones, 1995).

Kernel density derivative estimation have been of immerse significance in data exploration and visualization. The first and second derivatives also known as density gradient and density curvature are of vitally importance because auxiliary information not provided by examining the density function can be made available in kernel density derivative estimation. The secondary information provided by the estimation of the density derivatives enables the data analyst to make accurate decision and future prediction. Critical details about the composition of density function are not readily available with the examination of the density function alone but can only be obtained through the estimation of its derivatives. As a result of the usefulness of density derivative estimation, serious attention should be on the vital cases of density gradient estimation and density Hessian estimation which are the first and second derivatives before considering other higher derivative with more complex mathematical formulation (see Charnigo and Srinivasan, 2011; Henderson and Parmeter, 2015; Sasaki et al., 2015).

The application of kernel density estimator is largely hindered in practice by the complexity associated with the bandwidth selection procedure because the functionality of the estimator depends on the bandwidth. The problem of accurate bandwidth selection is more critical in higher order kernel estimation and kernel density derivative estimation because of the intricacy of their mathematical formulation that requires high computational know-how. The complication of bandwidth selection is more evident in higher dimensional setting since most application of kernel estimation is in multivariate case with different forms of parameterization. A superfluity of bandwidth selectors exists in literature for density estimation but little progress has been made in higher order kernel estimation and kernel density derivative estimation. In spite of the numerous bandwidth selectors, no particular method has been applied in all circumstances due to the variation and structural difference of the estimators. There are several bandwidth selectors for higher order kernel estimation and kernel density derivatives estimation in univariate case while the multivariate case is neglected due to the variations in their parameterizations which involve complex mathematical manipulation (see Härdle et al., 1990; Dobrovodov and Rud'ko, 2010; Chacón and Duong, 2013; Somé and Kokonendji, 2021).

The aim of this paper is to examine two methods of measuring the performance of kernel density estimator which are kernel density derivative estimation and higher order kernel estimation. These two methods are asymptotic mean integrated squared error (AMISE) reducing strategies either in the bias or variance components. The performance of a kernel method is a major determinant of the usability of the kernel method, hence the examination of these techniques. The results of the examination reveal that the kernel density derivative estimation is superior to the higher order kernel estimation in performance with the AMISE as the criterion function. The scope of this paper will be limited to the sixth order kernel estimator and the density curvature estimator because most benefits of higher order

kernels are in fourth order kernel while the density gradient estimator helps in the identification of modes of distributions. The paper primarily compares the performance of kernel density derivative and higher order kernel using the Gaussian kernel function due to the continuous differentiability of the distribution. The other part of this paper is organized as follows. In Section 2, the general form of the kernel estimator and its performance measure is presented while Sections 3 presents higher order kernel estimation and kernel density derivative estimation with the performance measure. Section 4 is the results and discussion using real data with emphasis on univariate and bivariate cases only. Section 5 concludes the paper.

2. Kernel Density Estimator.

The kernel estimator proposed by Rosenblatt (1956) and Parzen (1962) is a weighting function with its univariate form given as

$$\hat{f}(x) = \frac{1}{nh_x} \sum_{i=1}^n K\left(\frac{x - X_i}{h_x}\right), \tag{2.1}$$

where $K(\cdot)$ is the kernel function, n is the sample size, $h_x > 0$ is bandwidth (known as smoothing parameter), x represents the range of the observations and X_i are the observations. The kernel function is symmetric and non-negative function satisfying the following conditions

$$\int K(x)dx = 1, \int xK(x)dx = 0 \text{ and } \int x^2K(x)dx = \mu_2(K) \neq 0. \tag{2.2}$$

The conditions in (2.2) suggest that the kernel function is a probability density function because the function must integrate to one with a zero mean and a variance greater than zero (Scott, 2015). The choice of the kernel function is not critical owing to the fact that most kernel functions are probability density function but the choice of the bandwidths has been critically investigated with no single universally acceptable rule in all situations. The performance of the kernel estimator in (2.1) can be assessed using several error criteria functions but the asymptotic mean integrated squared error (AMISE) is consider in this paper due to its inclusion of dimension unlike other error criteria that are dimensionless. The inclusion of dimension in kernel density estimation is a characteristic that provides significant benefits in practical applications. The AMISE has two components which are the integrated variance and integrated squared bias given by

$$AMISE(\hat{f}(\cdot)) = \frac{R(K)}{nh_x} + \frac{1}{4}\mu_2(K)^2 h_x^4 R(f''). \tag{2.3}$$

In Equation (2.3), $R(K) = \int K(x)^2 dx$ is usually known as roughness of the kernel while $\mu_2(K)^2$ represents the kernel variance. The quantity $R(f'') = \int f''(x)^2 dx$ is the roughness of the unknown probability distribution for the estimation. The two components of the AMISE depend on the bandwidth which regulates their contributions to the AMISE respectively. The bandwidth with the minimum AMISE called the optimal bandwidth is of the form

$$h_{x-AMISE} = \left[\frac{R(K)}{\mu_2(K)^2 R(f'')} \right]^{1/(d+4)} \times n^{-1/(d+4)}, \tag{2.4}$$

where d is known as the dimension of the kernel estimator. The kernel estimator is mostly applied in the multivariate setting especially the two-dimensional kernel estimator that bridges the univariate and other higher dimensional kernel estimator. The bivariate kernel density estimation deals with two random variables jointly and the estimator is

$$\hat{f}(x, y) = \frac{1}{nh_x h_y} \sum_{i=1}^n K\left(\frac{x - X_i}{h_x}, \frac{y - Y_i}{h_y}\right), \tag{2.5}$$

where $h_x > 0$ and $h_y > 0$ are bandwidths for X and Y respectively while x and y are the ranges of the observations. The bivariate kernel estimator is an effective analytical tool for data analysis and visualization either as a wireframe or contour plot that usually reveal hidden information in the observations (Silverman, 2018; Siloko and Siloko, 2019). The product form of the bivariate kernel involves the multiplication of different univariate estimators given as

$$\hat{f}(x, y) = \frac{1}{nh_x h_y} \sum_{i=1}^n K\left(\frac{x - X_i}{h_x}\right) K\left(\frac{y - Y_i}{h_y}\right). \tag{2.6}$$

The product kernel estimator is commonly employed in density estimation and mostly beneficial when there are differences in the various axes of the data to be analyzed. Kernel estimation is reasonably practicable within low and moderate dimension since at higher dimension, data seems to be sparse and that will generate unstable estimate. Hence, the bivariate estimator is of wide applicability in density estimation due to easy accessibility of bivariate data (Scott, 2015). The AMISE of the bivariate product kernel estimator under the regular assumptions is given as

$$AMISE(\hat{f}(x, y)) = \frac{R(K)}{nh_x h_y} + \frac{h_x^4}{4} \mu_2(K)^2 \iint \left(\frac{\partial^2 f}{\partial x^2}\right)^2 dx dy + \frac{h_y^4}{4} \mu_2(K)^2 \iint \left(\frac{\partial^2 f}{\partial y^2}\right)^2 dx dy \quad (2.7)$$

The AMISE of the bivariate kernel estimator also require higher derivatives of the function to be estimated which needs approximation as in the univariate kernel estimator.

3. Higher Order Kernel and Kernel Density Derivative of Gaussian Estimator.

Higher-order kernel estimation and kernel density derivative estimation are AMISE reduction techniques that use large bandwidths. The Gaussian kernel function is our focus because the function is continuously differentiable and with the established evidence of its extensive applicability in data mining and other related statistical fields.

3.1 Higher Order Kernel.

The rationale behind the application of higher order kernels is mainly due to their fast convergence rates with their bias reduction property as against the lower order kernel estimators whose rates of convergence is slow, hence the application of higher order kernel estimators (Jones and Foster, 1993; Marron, 1994; Jones and Signorini, 1997; Ishiekwene and Osemwenkhae, 2006; Siloko et al., 2019a). A kernel function is usually of higher-order when the order denoted by p is greater than two, that is $p > 2$. Higher order kernels are usually constructed from their corresponding lower order kernels using several techniques such as the additive and multiplicative methods. Higher order kernels must satisfy the following conditions

$$\int K(x)dx = 1, \int x^i K(x)dx = 0, i = 1, 3, \dots, p - 1 \text{ and } \int x^p K(x)dx \neq 0, \quad (3.1.1)$$

where p is the order of the kernel which is the nonzero moment whereas the odd moments of any kernels are zero. The AMISE of the p^{th} order kernel density function is given by

$$AMISE^*(\hat{f}(\cdot)) = \frac{R(K)}{nh_x} + \frac{1}{(p!)^2} \mu_p(K)^2 h_x^{2p} R(f^{(p)}), \quad (3.1.2)$$

where $\mu_p(K)^2$ is the p^{th} moment of the kernel. The bandwidth that minimizes the AMISE is

$$h_{x-AMISE}^* = \left[\frac{(p!)^2 R(K)}{2p \mu_2(K)^2 R(f^{(p)})} \right]^{1/(2p+1)} \times n^{-1/(2p+1)}. \quad (3.1.3)$$

Assuming in the bandwidth that minimizes the AMISE of higher-order kernel estimation, the unknown probability density function is the Normal distribution and the Gaussian kernel is employed, then the roughness of the Gaussian kernel function which is denoted by $R(f^{(p)})$ is given by

$$R(f^{(p)}) = \int \{f^{(p)}(x)\} dx = \frac{1}{2\pi\sigma^{(2p+1)}} \Gamma\left(\frac{2p+1}{2}\right), \quad (3.1.4)$$

where $\Gamma(\cdot)$ is the gamma function and $p = 2, 4, 6, \dots, \infty$ is the order of the kernel function. The optimal bandwidth that produces the least AMISE value is given by

$$AMISE^*(\hat{f}(\cdot)) = \left[\frac{R(K)^{2p} R(f^{(p)})}{\mu_p(K)^{-2}} \right]^{1/(2p+1)} \times n^{-2p/(2p+1)}. \quad (3.1.5)$$

The optimal bandwidth of higher order kernel estimation is of order $O(n^{-1/(2p+1)})$ while the order of its AMISE is $O(n^{-2p/(2p+1)})$. The large bandwidths necessary for the implementation of higher-order kernels is occasioned by the order of the bandwidth. Higher-order kernel estimation also require considerable large sample size in harnessing their potential benefits (Siloko et al., 2019b).

3.2 Kernel Density Derivative Estimation.

The derivative of a kernel function can be obtained by differentiating the kernel estimator. Given a kernel function K that is sufficiently differentiable r times, then the r^{th} kernel derivative of Equation (2.1) is of the form

$$\hat{f}^{(r)}(x) = \frac{1}{nh_x^{r+1}} \sum_{i=1}^n K^{(r)}\left(\frac{x - X_i}{h_x}\right), \quad (3.2.1)$$

where $K^{(r)}$ is the r^{th} derivative of K which is usually a symmetric probability density function. Again, the AMISE of the derivative of the kernel estimator if K is continuously differentiable is

$$AMISE(\hat{f}^{(r)}(\cdot)) = \frac{R(K^{(r)})}{nh_x^{2r+1}} + \frac{1}{4}h_x^4\mu_2(K)^2R(f^{(r+2)}), \tag{3.2.2}$$

where $R(K^{(r)})$ is roughness of the r th kernel derivative estimator, $\mu_2(K)^2$ is variance and $R(f^{(r+2)})$ is roughness of r th unknown probability density function. Every derivative order in kernel density derivative estimation attracts two supplemental powers in the asymptotic variance of the AMISE while the order of the asymptotic squared bias is constant (Scott, 2015). The bandwidth with minimum AMISE value in (3.2.2) is

$$h_{x-AMISE}^r = \left[\frac{(2r+1)R(K^{(r)})}{\mu_2(K)^2R(f^{(r+2)})} \right]^{1/(2r+5)} \times n^{-1/(2r+5)}. \tag{3.2.3}$$

Again, the r th roughness of the Gaussian kernel denoted by $R(K_\phi^{(r)})$ is given as

$$R(K_\phi^{(r)}) = \frac{(2r-1)!!}{2^{r+1}\sqrt{\pi}}. \tag{3.2.4}$$

If the unknown distribution is the Gaussian distribution with mean (μ) is zero and variance is σ^2 , then the r th roughness with respect to the derivative of the distribution is given as

$$R(f^{(r+2)}) = \left\{ \frac{2^{(2r+5)}(r+2)!\sqrt{\pi}}{(2r+4)!\sigma^{-(2r+5)}} \right\}^{-1} \tag{3.2.5}$$

Substituting Equation (3.2.5) into Equation (3.2.3) and expressing it in terms of dimension, we have the bandwidth with the minimum AMISE of the kernel density derivative estimator as

$$h_{x-AMISE}^r = \left[\frac{2^{(d+2r+4)}(r+2)!\sqrt{\pi}(2r+1)R(K^{(r)})}{(2r+4)!\mu_2(K)^2\sigma^{-(d+2r+4)}} \right]^{1/(d+2r+4)} \times n^{-1/(d+2r+4)}, \tag{3.2.6}$$

where σ is usually the standard deviation of the observations. The bandwidth for kernel density derivative estimation must be appropriately chosen because quality density estimator often times does not give rise to superior kernel density derivatives especially as the order of the derivative increases (Sasaki, 2015). The least AMISE value of the optimal bandwidth in (3.2.6) is

$$AMISE(\hat{f}^{(r)}(\cdot)) = \left(\frac{2r+5}{4} \right) R(K^{(r)})^{4/(2r+5)} \left[\frac{\mu_2(K)^2 R(f^{(r+2)})}{2r+1} \right]^{2r+1/(2r+5)} n^{-4/2r+5}. \tag{3.2.7}$$

The derivative of a function is more boisterous than the function itself, hence large bandwidths are usually required for kernel density derivative estimation. In the first and second derivatives, the bandwidths with the minimal AMISE values are of orders $O(n^{-1/7})$ and $O(n^{-1/9})$ with the AMISE having orders $O(n^{-4/7})$ and $O(n^{-4/9})$ respectively.

3.3 The Gaussian Kernel Estimator.

The Gaussian kernel is one of the popular kernel estimators because of its production of smooth density estimates and possession of derivatives of all orders which promoted its extensive application in kernel density derivative estimation. The classical second order Gaussian kernel estimator is the limiting case of the beta polynomial kernel family whose univariate form is given by

$$K_{2,\phi}(x) = \frac{1}{\sqrt{2\pi}} \exp - \left(\frac{x^2}{2} \right). \tag{3.3.1}$$

The corresponding bivariate form of the Gaussian kernel estimator is given as

$$K_{2,\phi}(x, y) = \frac{1}{2\pi} \exp - \left(\frac{x^2 + y^2}{2} \right). \tag{3.3.2}$$

The fourth and sixth order kernel functions of the Gaussian kernel estimator are

$$K_{4,\phi}(x) = \frac{1}{2\sqrt{2\pi}} (3 - x^2) \exp - \left(\frac{x^2}{2} \right). \tag{3.3.3}$$

$$K_{6,\phi}(x) = \frac{1}{8\sqrt{2\pi}} (15 - 10x^2 + x^4) \exp - \left(\frac{x^2}{2} \right). \tag{3.3.4}$$

The derivatives of the Gaussian kernel are usually deduced from the Hermite polynomial family. The Gaussian kernel derivative is denoted by $K^{(r)}(x) = (-1)^r H_r(x)K(x)$ where $r = 0, 1, 2, \dots$ is the derivative order and $H_r(x)$ is the r th Hermite polynomial. The Hermite polynomials have the following values for its first six members: $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$, $H_4(x) = x^4 - 6x^2 + 3$, and $H_5(x) = x^5 - 10x^3 + 15x$. The usual estimator of the Gaussian kernel density derivative is of the form

$$\hat{f}^{(r)}(x) = \frac{(-1)^r}{\sqrt{2\pi n} h_x^{r+1}} \sum_{i=1}^n H_r(x) \exp\left[-\frac{1}{2} \left(\frac{x-X_i}{h_x}\right)^2\right] \tag{3.3.5}$$

The estimator in Equation (3.3.5) is the generalized form of estimating the r th kernel density derivative of the popular Gaussian distribution. If $r = 0$, the resulting kernel estimator is the traditional second order kernel estimator in Equation (2.1). Another widely used kernel function of the beta polynomial family is the Epanechnikov kernel also known as the optimal kernel with respect to the asymptotic mean integrated squared error (Siloko et al., 2019b).

4. Result and Discussions.

The performance of higher-order kernel estimation and kernel density derivative estimation is investigated with the AMISE as the criterion function owing to its mathematical tractability. The graphical analysis and computational results obtained from the two methods were with the aid of Mathematica version 12.3 software (Wolfram Research, Inc.). Higher-order kernel estimation and kernel density derivative estimation usually employ large bandwidths in their implementation as the order increases and that accounted for their bias and variance reduction property that often translates to AMISE reduction. A sample size of 1000 and two real data sets were analyzed using the Gaussian kernel estimator for the higher-order kernel estimation and kernel density derivative estimation.

The results in Table 1 are that of a randomly generated data of sample size of 1000 with standard deviation of 23.2069. The standard deviation is one of the statistical parameters required in the computation of bandwidths for higher order kernel estimation and kernel density derivative estimation. The performance of higher order kernel and kernel density derivative estimations depend on the size of the bandwidth and both methods require large bandwidth for their implementation with increase in the kernel order and derivative order respectively. The bandwidths of higher order kernel estimation are larger than the bandwidths of kernel density derivative estimation but the AMISE values of kernel density derivative estimation are smaller which is an indication of better performance. Again, as seen in Table 1, the benefit of higher order kernel estimation is mainly in the fourth order and reduction in the AMISE value in subsequent orders tend to be minimal (Ishiekwene and Osemwenkhae, 2006). In kernel density derivative estimation, the benefits are mainly in the first and second derivative estimations and graphical visualization of data beyond the second derivative order oftentimes may be difficult especially with higher dimensions. Although, the bandwidths of higher order kernel estimation are larger than the bandwidths of kernel density derivative estimation but with smaller AMISE values and this is occasioned by the two-extra power of the bandwidth in the variance component.

Table 1. Kernel Order, Derivative Order, Bandwidths and AMISE of Simulated Data of Size n=1000

Kernel Order (p)	Bandwidth	AMISE	Derivative Order (r)	Bandwidth	AMISE
2	6.174541	0.00005710845058	0	6.174541	0.00005710845058
4	14.05622	0.00002257767253	1	8.379175	0.00000041956514
6	20.95916	0.00001458086719	2	10.12230	0.00000000447963
8	26.96567	0.00001111508580	3	11.50464	0.00000000005453
10	32.29093	0.00000917284105	4	12.61981	0.00000000000074

The first data set is the Annual Snowfall in Buffalo with a sample size of 63 observations (Scott, 2015). Buffalo is one of the largest cities in the state of New York with a record of an annual snowfall of 84.8 inches which implies that the city is snowier than other areas in the state. The kernel estimates of the snowfall data show that the data are unimodal which is within the region of the annual snowfall. Figure 1 is the kernel estimate of the second order kernel and fourth order kernel while Figure 2 is the estimate of the sixth order kernel with the second order kernel to sixth order kernel estimates. The kernel estimates of the zeroth derivative and first derivative is in Figure 3 while Figure 4 is the kernel estimate of the second derivative with the zeroth to second derivative estimates. The kernel estimates of the snowfall data with the higher order kernel displayed unimodality but with the first derivative estimation, the data are presented to be bimodal and this is attributed to the fact that obscured features in the data set in kernel density estimation can be clearly revealed with the kernel density derivative estimation particularly the density gradient.

The results in Table 2 and Table 3 are for the two AMISE reduction techniques using the univariate Gaussian kernel function. Generally, both techniques require large bandwidths for their effective implementation but the size of the bandwidths of higher order kernel estimation is larger in comparison with the size of bandwidths of kernel density derivative estimation. Again, the AMISE of kernel density derivative estimation is smaller than the AMISE of higher order kernel estimation and that authenticates the fact that kernel density derivative estimation outperform higher order kernel estimation. The estimate of the second order kernel and the zeroth derivative estimate are the same hence; the same bandwidths and AMISE values as seen in Table 2 and Table 3. The estimates of the second, fourth and the sixth order kernels in Figure 2(b) displayed similarity graphically; however, there is great variation in performance with reference to their AMISE values. As order of kernel increases, there is a reduction in the AMISE which is occasioned by the size of the bandwidths and the order of the kernel but there is the tendency of smoothing away some beneficial statistical features in the observations being estimated.

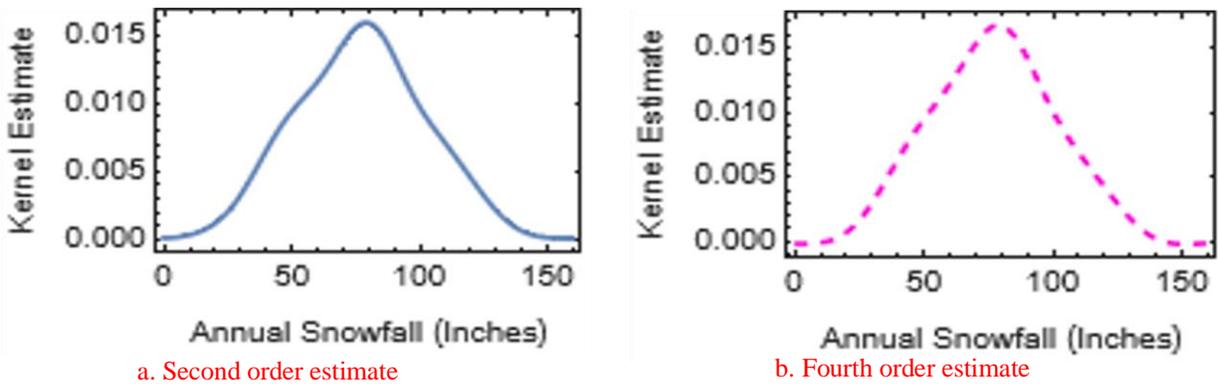


Figure 1: Second order kernel estimate and fourth order kernel estimate of the snowfall data.

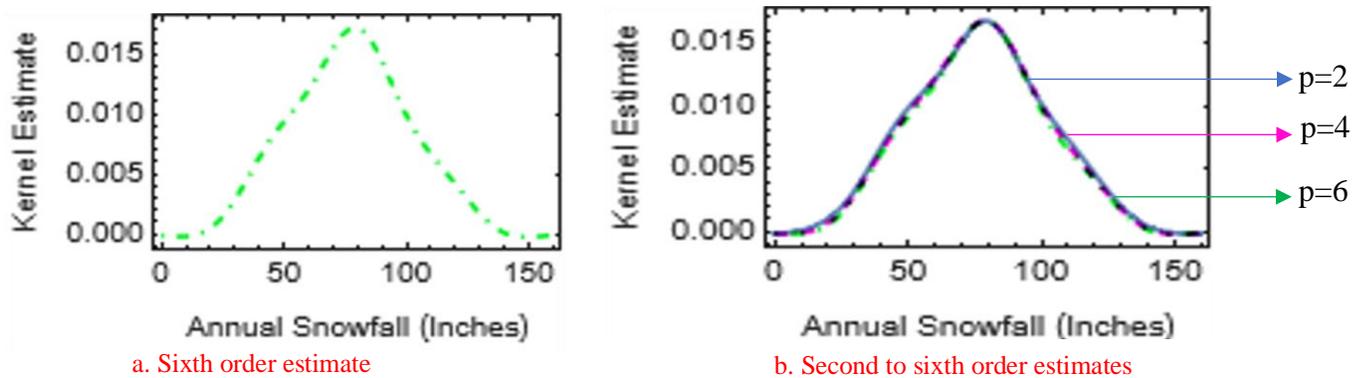


Figure 2: Sixth order kernel estimate and second order to sixth order kernel estimates of the snowfall data.

Table 2. Order, Bandwidths, Bias², Variance and AMISE of Higher Order Kernel Estimation for First Data.

Kernel Order	Bandwidth (h_x)	Bias ²	Variance	AMISE
2	10.733387	0.000104293624	0.000417174496	0.000521468120
4	19.110631	0.000029287986	0.000234303886	0.000263591872
6	25.925804	0.000014392659	0.000172711907	0.000187104566
8	31.727672	0.000008820563	0.000141129013	0.000149949576
10	36.834498	0.000006078127	0.000121562538	0.000127640665

Table 3. Order, Bandwidths, Bias², Variance and AMISE of Kernel Density Derivation Estimation for First Data.

Derivative Order	Bandwidth (h_x)	Bias ²	Variance	AMISE
0	10.733387	0.000104293624	0.000417174496	0.000521468120
1	12.437245	0.00000872797	0.000001163729	0.000002036526
2	13.762146	0.00000008503	0.00000006803	0.00000015306
3	14.791899	0.00000000095	0.00000000054	0.00000000149

Generally, data that exhibits multimodality when estimated with large bandwidths in kernel estimation may tend to display unimodality but with a reduction in the value of the AMISE in terms of performance. Despite the fact that the error criterion is the determinant of the acceptability of a method in kernel density estimation, efforts had been tailored statistically towards the retention of inherent characteristics of the data for prediction and decision making. The determination of vital statistical features in kernel estimation depends greatly on the magnitude of the bandwidth. The results in Table 2 and Table 3 are the performances of the methods considered and from the results; the reduction in AMISE value is more noticeable with the kernel density derivative estimation owing to the fact that its variance component requires two additional powers in the bandwidths. Hence, there are reductions in the variance and bias terms that translate to reduction in the AMISE unlike the higher order kernel estimation whose reduction is mainly in the bias term of the AMISE. Theoretically, in kernel density derivatives estimation, any order of interest can be estimated but the benefits are mainly in the gradient and hessian estimation with little or no benefits in other higher derivatives order. Regarding higher-order kernel estimation with similar estimates, the benefit is mainly in the fourth order kernel and reductions in further higher order may be minimal.

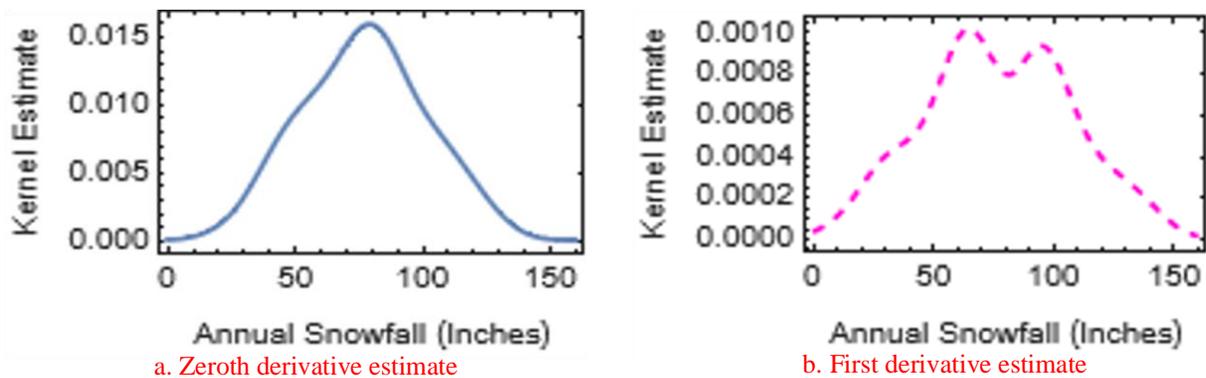


Figure 3: Zeroth derivative estimate and first derivative estimate of the snowfall data.

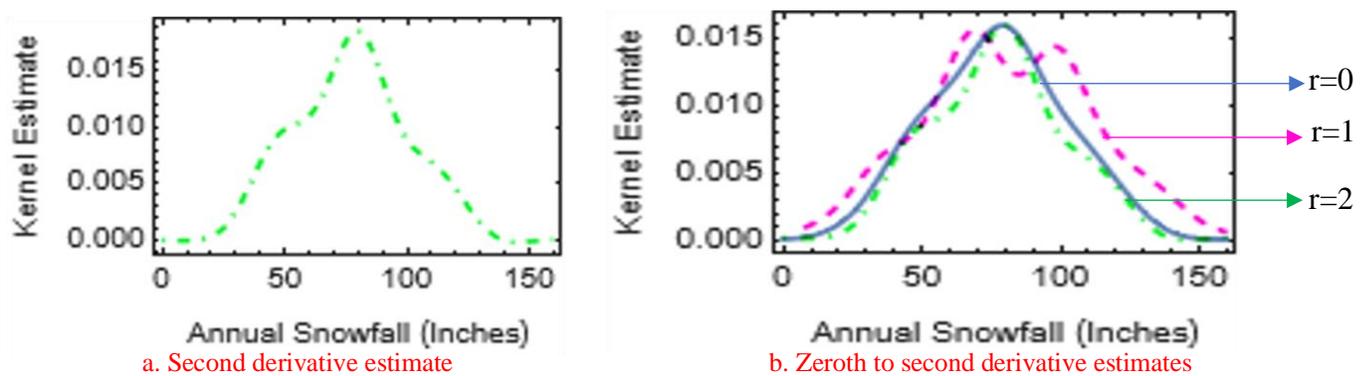


Figure 4: Second derivative estimate and zeroth to second derivative estimates of the snowfall data.

The second data investigated is the old faithful data which comprises of the duration of eruption and waiting time betwixt eruptions of the geyser situated in Yellowstone National Park in United States of America (Azzalini and Bowman, 1990). The old faithful data are made up of 272 data points for the two axes respectively and the bivariate

kernel estimates are bimodal indicating that eruption times and the time before the immediate next eruption often demonstrates a distribution that is usually bimodal. The bivariate kernel estimates of the second, fourth and sixth order kernels are in Figure 5, Figure 6 and Figure 7 with the bimodality been obviously displayed. The fundamental role of kernel density estimation is data exploratory analysis and visualization due to its ability of highlighting significant features in the data unlike other data visualization tools like the scatterplots where the observer is only drawn to the data cloud while inherent features of the data are hidden. Again, the contour plots of the bivariate fourth order kernel estimate and sixth order kernel estimate in Figure 6 and Figure 7 revealed the data point that are regarded as outliers while maintaining the bimodal nature of the data set.

The bimodality of the bivariate kernel estimates has evidenced that eruption times and time intermission prior to the next eruption is distinctly correlated. The bivariate kernel estimates of the density gradient estimation and density curvature estimation are in Figure 8 and Figure 9 with all the estimates depicting the bimodal nature of the data. The bivariate estimate of the zeroth derivative estimation is same as the bivariate estimate of the second order kernel because the zeroth derivative estimation is same as the second order kernel estimation. Again, the bivariate bandwidths of the kernel density derivative estimation are smaller than the higher order kernel estimation but with smaller AMISE values as in the univariate case.

The application of kernel density derivative estimation and higher-order kernel estimation becomes very complex with increase in the dimensions of the observations. The complexity of estimation of higher dimensions is more noticeable in the graphical presentation of the observation because for higher dimensions above four-dimensional forms, visualization of observations which is very important in kernel density estimation becomes practically difficult. This difficulty connected with higher dimension in nonparametric estimation is known as the curse of dimensionality effects and that has limited kernel density estimation to the bivariate case; although numerical computation can be extended to higher dimensions (Scott, 2015; Siloko et al., 2020b).

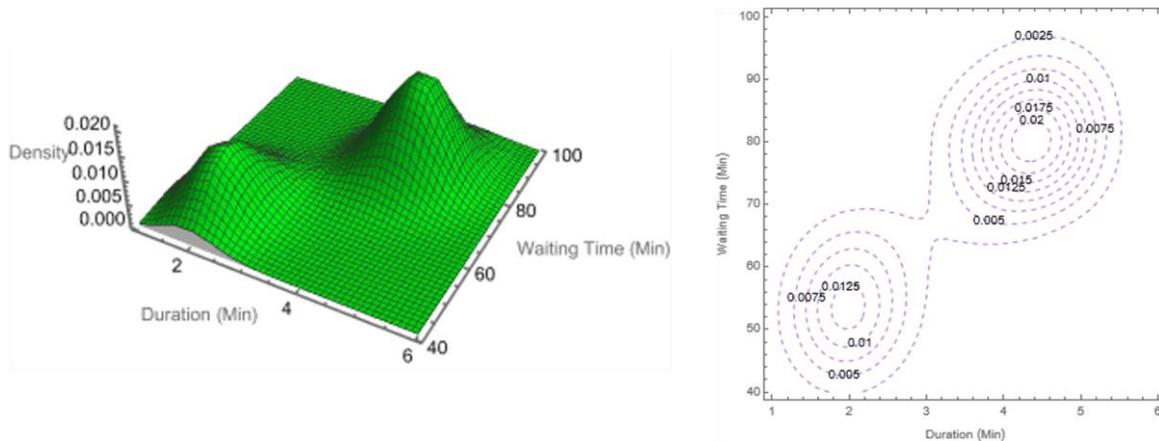


Figure 5: Surface and contour plots of bivariate second order estimates of the old faithful data.

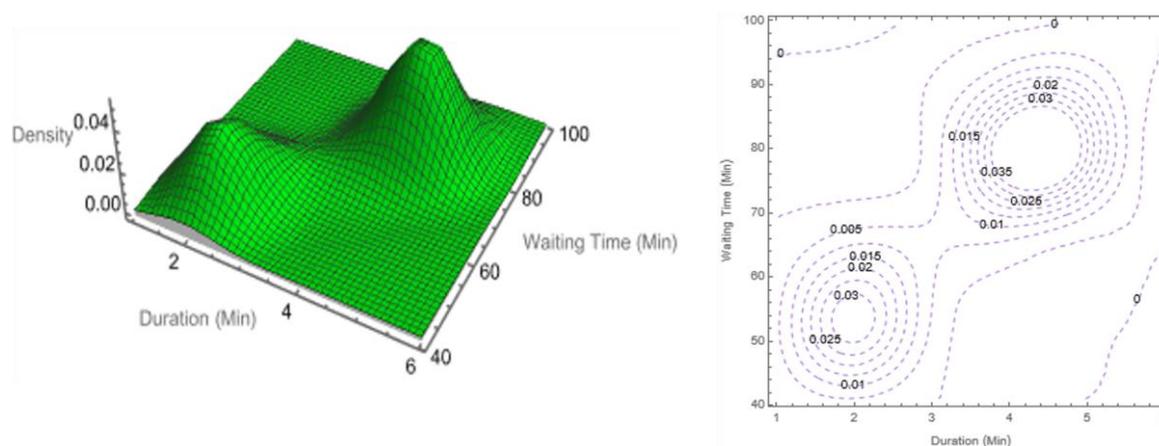


Figure 6: Surface and contour plots of bivariate fourth order estimates of the old faithful data.

Table 4. Order, Bandwidths, Bias², Variance and AMISE of Higher Order Kernel Estimation for Second Data.

Kernel Order	Bandwidth (h_x)	Bandwidth (h_y)	Bias ²	Variance	AMISE
2	0.44851327	5.34093014	0.000483333604	0.000122131879	0.000605465481
4	0.64816082	6.02476642	0.000052025594	0.000074920081	0.000126945675
6	0.76027763	7.58521646	0.000005034085	0.000050731890	0.000055765975

Table 5. Order, Bandwidths, Bias², Variance and AMISE of Kernel Density Derivation Estimation for Second Data.

Derivative Order	Bandwidth (h_x)	Bandwidth (h_y)	Bias ²	Variance	AMISE
0	0.44851327	5.34093014	0.000483333604	0.000122131879	0.000605465481
1	0.53852301	6.41277291	0.000000191643	0.000001775869	0.000001967512
2	0.60810265	7.24133253	0.000000070265	0.000000099396	0.000000169661

The bias and variance reduction property of the univariate kernel density derivative estimation is also demonstrated in the bivariate estimation with smaller AMISE values as presented in Table 5 in comparison with Table 4. Again, with kernel density derivative estimation after the second derivative order, there is usually little or no gain regarding reduction in the AMISE since the values will be too minimal because of the concurrent reduction in the components of the AMISE. Kernel density derivative estimation and higher-order kernel estimation have demonstrated the potential of reducing the AMISE either in the bias component or both variance and bias terms concurrently depending mainly on the magnitude of the bandwidths and other parameters of interest such as the roughness of the kernel function and roughness of the distribution.

Generally, in kernel density estimation, the superiority or dominance of one method over other existing methods can be ascertained by the value of a known performance measure such as the AMISE or other criteria functions (Jarnicka, 2009). In spite of the establishment of both methods using large bandwidths in their implementations, the inherent statistical features of the observations investigated with these methods were retained as seen in the kernel estimates. Retention of statistical characteristics of observations is the central point of kernel density estimation especially in data visualization. However; the results vividly demonstrated numerically and graphically the dominance of the kernel density derivative estimation over the higher-order kernel estimation in terms of performance.

5. Conclusions.

Higher-order kernel estimation and kernel density derivative estimation are AMISE reduction techniques that employ large bandwidths for their implementation as the order increases. Although both methods are AMISE reduction techniques; this paper tends to ascertain their AMISE reduction capacities which establishes the superiority of one of the techniques over the other. This study undoubtedly revealed that higher-order kernel estimation is a bias reducing approach that translates to AMISE reduction while kernel density derivative estimation is a bias and variance reduction strategy that resulted in the reduction of AMISE due to the extra-two powers of the bandwidth usually associated with the variance term for every derivative order. The requirements of large bandwidths for both methods are hinged on the noisy nature connected with higher-order kernel estimation and the kernel density derivative estimation of kernel functions in comparison with lower order kernels and the function itself. Numerical evaluation of the performance of kernel density derivative estimation and higher-order kernel estimation using the AMISE shows that the former outperformed the later. Again, in identification of statistical features in the analyzed data, kernel density derivative estimation particularly the density gradient can reveal hidden features that could be of help in prediction and decision-making process. Hence; kernel density derivative estimation which is a bias and variance reduction technique is highly recommended in nonparametric density estimation particularly for analyzing data with unknown information.

Conflicts of Interest

The authors declared that they have no conflict of interest.

Acknowledgement

The authors sincerely appreciate the anonymous reviewers for conscientiously going through the manuscript and making positive recommendations that improved the manuscript.

References

1. Azzalini, A. and Bowman, A. W. (1990). A look at some data on the old faithful geyser. *Applied Statistics*, 39, 357–365.
2. Bonnier, A., Finne, 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.
3. Chacón, J. E. & Duong, T. (2013). Data-driven density derivative estimation with applications to nonparametric clustering and bump hunting. *Electronic Journal of Statistics*, 7, 499–532.
4. Charnigo, R. & Srinivasan, C. (2011). Self-consistent estimation of mean response functions and their derivatives. *Canadian Journal of Statistics*, 39(2), 280–299
5. Dobrovidov, A.V. & Rud'ko, I.M. (2010). Bandwidth selection in nonparametric estimator of density derivatives by smoothed cross validation method. *Autom. Remote Control*, 71, 209–224.
6. Green, P. J., Latuszyski, K., Pereyra, M. & Robert, C. P. (2015). Bayesian Computation: A summary of the current state, and samples backwards and forwards. *Statistics and Computing*, 25, 835–862.
7. Härdle, W., Marron, J. S. & Wand, M. P. (1990). Bandwidth choice for density derivatives. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 52, 223–232.
8. Helu, A., Samawi, H., Rochani, H., Yin, J. & Vogel, R. (2020). Kernel density estimation based on progressive type-II censoring. *Journal of Korean Statistical Society*, 49, 475–498.
9. Henderson, D. J. & Parmeter, C. F. (2015). *Applied nonparametric econometrics*, Cambridge University Press.
10. Ishiekwene, C. C. & Osemwenkhae, J. E. (2006). A comparison of forth-order window width selectors in kernel density estimation (A univariate case). *ABACUS*, 33, 14–20.
11. Jarnicka, J. (2009). Multivariate kernel density estimation with a parametric support. *Opuscula Mathematica*, 29(1), 41–45.
12. Johnson, J. E., Laparra, V., Pérez-Suay, A., Mahecha, M. D. & Camps-Valls, G. (2021). Kernel methods and their derivatives: Concept and perspectives for the Earth System Sciences, <https://doi.org/10.1371/journal.pone.0246775>.
13. Jones, M. C., Foster, P. J. (1993). Generalized jackknifing and higher order kernels. *Journal of Nonparametric Statistics*, 3, 81–94.
14. Jones, M. C. (1992). Potential for automatic bandwidth choice in variations on kernel density estimation. *Statistics and Probability Letter*, 13, 351–356.
15. Jones, M. C. & Signorini, D. F. (1997). A comparison of higher-order bias kernel density estimators, *Journal of American Statistical Association*, 92, 1063–1073.
16. King, T. L., Bentley, R. J., Thornton, L. E. & Kavanagh, A. M. (2016). Using kernel density estimation to understand the influence of neighborhood destinations on BMI, *BMJ*, <https://doi: 10.1136/bmjopen-2015-008878>.
17. Li, C. L., Chang, W. C., Miroueh, Y., Yang, Y. & Poczos, B. (2019). Implicit kernel learning, *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, Naha, Okinawa, Japan, PMLR Vol. 89.
18. Marron, J. S. (1994). Visual understanding of higher-order kernels. *Journal of Computational and Graphical Statistics*, 3(4), 447–458.
19. Marron, J. S. & Wand, M. P. (1992). Exact mean integrated squared error. *Annals of Statistics*, 20, 712–736.
20. Nan, Y. & Ji, H. (2020). Deep learning for handling kernel/model uncertainty in image deconvolution. *Proceedings of the IEEE on Computer Vision and Pattern Recognition (CVPR)*, pp. 2388–2397.
21. Parzen, E. (1962). On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics*, 33, 1065–1076.
22. Raykar, V. Duraiswami, C. R. & Zhao, L. H. (2015). Fast computation of kernel estimators, *Journal of Computational and Graphical Statistics*, 19(1), 205–220.
23. Rojas, F. (2017). An application of kernel estimation method to the demand in two-stage inventory model. *Cogent Engineering*, <https://doi.org/10.1080/23311916.2017.1362803>
24. Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27, 832–837.

25. Sasaki, H., Noh, Y. & Sugiyama, M. (2015). Direct density derivative estimation and its application in KL-Divergence approximation. Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS), San Diego, CA, USA. JMLR: W&CP, Volume 38.
26. Scott, D. W. (2015). Multivariate density estimation: Theory, practice and visualization, Second edition, Wiley, New Jersey.
27. Siloko, I. U., Ikpotoke, O. & Siloko, E. A. (2019a). On hybridizations of fourth order kernel of the beta polynomial family. Pak. J. Stat. Oper. Res., 15(3), 819–829.
28. Siloko, I. U., Ikpotoke, O. & Siloko, E. A. (2019b). A key note on performance of smoothing parameterizations in kernel density estimation. Tanzania Journal of Science, 45(1), 1–8
29. Siloko, I. U., Ishiekwene, C.C. & Oyegue, F.O. (2018). New gradient methods for bandwidth selection in bivariate kernel density estimation. Mathematics and Statistics, 6(1), 1–8.
30. Siloko, I. U., Ukhurebor, K. E., Siloko, E. A., Enoyoze, E., Bobadoye, A. O., Ishiekwene C. C., Uddin, O. O., & Nwankwo, W. (2021). Effects of some meteorological variables on cassava production in Edo State, Nigeria via density estimation. *Scientific African*, e00852, <https://doi.org/10.1016/j.sciaf.2021.e00852>
31. Siloko, I. U., Nwankwo, W. & Siloko, E. A. (2020a). A new family of kernels from the beta polynomial kernel with applications in density estimation. International Journal of Advances in Intelligent Informatics, 6(3), 235–245.
32. Siloko, I. U., Siloko, E. A. & Ikpotoke, O. (2020b). A mini review of dimensional effects on asymptotic mean integrated squared error and efficiencies of selected beta kernels. Jordan Journal of Mathematics and Statistics, 13(3), 327–340.
33. Siloko, I. U. & Siloko, E. A. (2019), Density estimation and data analysis using the kernel method, Proceedings of 56th Annual National Conference (Mathematics Science) of Mathematical Association of Nigeria, 218–226
34. Silverman, B. W. (2018). Density estimation for statistics and data analysis, Routledge, Abingdon, UK
35. Somé, S. M. & Kokonendji, C. C. (2021). Bayesian selector of adaptive bandwidth for multivariate gamma kernel estimator on $[0, \infty)^d$, Journal of Applied Statistics, <https://doi.org/10.1080/02664763.2021.1881456>
36. Spencer, C. J., Yakymchuk, C. & Ghaznavi, M. (2017). Visualizing data distributions with kernel density estimation and reduced Chi-squared statistics. Geoscience Frontiers, 8(6), 1247–1252.
37. Stanislaw, W. (2018). Kernel density estimation and its application, ITM Web of Conferences, XLVIII Seminar of Applied Mathematics, 23, 00037.
38. Wand, M. P. & Jones, M. C. (1995). Kernel Smoothing, Chapman and Hall, London.