Hacettepe Journal of
Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 54 (3) (2025), 1094 – 1106
DOI : 10.15672/hujms.1605499

Research Article

Optimizing bandwidth parameter estimation for non-parametric regression using fixed-form threshold with Dmey and Coiflet wavelets

Delshad Shaker Ismael Botani*, Nazeera Sedeek Kareem, Taha Hussein Ali, Bekhal Samad Sedeeq

Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil, Iraq

Abstract

The article aims to reduce the effect of data noise or outliers and estimate the optimal bandwidth parameter used in nonparametric regression models using a proposed method based on wavelet analysis, specifically Dmey and Coiflet wavelets with fixed-form threshold and apply the soft threshold, particularly when the data have long-tailed and multimodal distributions (abnormal distribution). The fixed-form threshold level value estimates the bandwidth instead of the classical method (geometric, arithmetic mean, range, and median). A simulation study was used to examine the suggested method, comparing it with four other Nadaraya-Watson kernel estimators (classical techniques), using a MATLAB language created especially for this purpose with actual data. The findings show that the suggested method outperforms classical methods for all cases of simulations and real data in accurately estimating the bandwidth parameter of the non-parametric regression kernel function based on the mean square error criterion.

Mathematics Subject Classification (2020). 62G05, 62G07, 62G08

Keywords. Non-parametric regression, kernel estimator, bandwidth parameter, fixed-form threshold, wavelets

1. Introduction

The statistical technique of non-parametric regression is used to estimate the connection between variables without making assumptions about the relationship's functional form. It enables flexible modeling of complicated relationships and is frequently employed when the underlying data do not correspond to a particular parametric model. The accurate analysis of the unknown response function f is the objective of the regression analysis, given n data points $\{(x_i, y_i)\}_{i=1}^{n}$. The following is an appropriate model for the relationship [3]:

$$y_i = f(x_i) + \varepsilon_i; \quad i = 1, 2, \ldots, n. \tag{1.1}$$

*Corresponding Author.
Email addresses: delshd.botani@su.edu.krd (D. S. I. Botani), nazeera.kareem@su.edu.krd (N. S. Kareem), taha.ali@su.edu.krd (T. H. Ali), bikhal.sedeeq@su.edu.krd (B. S. Sedeeq)
Received: 25.12.2024; Accepted: 24.04.2025

Here, $f(x_i)$ represents an unidentified regression function, ε_i are random errors with constant variance σ_ε^2 and a mean of zero. The weighted average of the dependent variable is used in non-parametric regression, with the weights representing the distance between the independent variable data as specified by a smoothing parameter. The Nadaraya-Watson (NW) estimator of kernel function is a method employed for the estimation of non-parametric regression. By using a weighting technique, Nearby data points are given weights based on how near the predicted place they are. It determines the form and range of the influence of the data points on the calculated function. Kernel functions that are commonly utilized include Gaussian, Epanechnikov, and uniform kernels. The NW estimator was introduced by [19] and [24] as a non-linear approximation to a non-parametric regression model (NRM) that is dependent on empirical data. Observations at x_i close to x should provide information about the value of f at x if $f(.)$ is thought to be smooth [12]. Consequently, an estimate $f(x)$ should be able to be created using something similar to a local average of the data close to x.

The smoothing bandwidth parameter (h), which regulates how smooth the predicted function is, is a need for the NW kernel estimator. A smoother estimate is produced by a bigger bandwidth; a more detailed estimate, which may be constant or changeable, is produced by a lower bandwidth [21, 23].

The ideal bandwidth h of the NW kernel estimator is the value at which the mean integrated squared error (MISE) is minimized. Several methods are used to calculate the bandwidth value. When the bandwidth is unknown, Friedman and Stuetzle [13] employed robust reparability to determine the NRM's parts. Furthermore, a consistent and asymptotically normal estimator based on kernels was presented. Since non-parametric estimating relies on the geometric mean, Läuter [17] proposed a modification for the kernel function estimator that varied the bandwidth and decreased the bias more than the fixed h estimator. More information on bandwidth estimation techniques was provided by [23], while Härdle and Kelly [15] examined several techniques for estimating bandwidth that use least-squares cross-validation as an unbiased estimator. As an alternative to the geometric mean in the NW kernel estimator, the arithmetic mean $(\bar{x})$ was suggested by [11] as an improvement to the Adaptive Nadaraya-Watson (ANW) approach. According to [18], the ANW method should be modified based on smoothing parameter selection in kernel non-parametric regression using a bat optimization algorithm. Aljuhani and Al turk [7] proposed a new modification that relies on a range of the density function of the kernel for the ANW method instead of using the arithmetic mean. The findings supported earlier research and showed that the variable bandwidth NW kernel estimator performs better than the fixed NW kernel estimator. To alter the ANW approach, Ali and Qadir [5] proposed employing the robust mean (R-M), median (Me), and harmonic mean (H-M) of the kernel estimator. Using a universal threshold level value for the Daubechies wavelet coefficients of the kernel density function in the kernel estimator, Ali et al. [4] proposed to modify the ANW technique.

Dhafir et al. [20] discussed the NW estimator, which is crucial in smoothing techniques to estimate regression functions. It provides a foundation for understanding fixed and variable bandwidth methods in nonparametric regression, which is relevant to your study. Hassan and Hmood [14] introduced work that applies kernel smoothers and wavelets in the context of financial data, specifically to estimate stock return rates. The methodology of this study could be aligned with your approach to examining variable bandwidths in nonparametric smoothing, providing additional insights into practical applications.

Wavelets are highly regarded in signal processing and statistical estimation due to their ability to capture local features in data at various scales, making them ideal for nonparametric regression and density estimation. In this context, wavelets can provide a robust mechanism for bandwidth selection.

The non-parametric regression estimator known as the NW kernel is improved upon in this study. With a soft threshold, this improvement efficiently reduces the influence of data noise by determining the bandwidth based on the fixed-form threshold level value with (Dmey and Coiflets) wavelets of the kernel function. A thresholding technique is used for wavelet coefficients, and it is a value that determines which coefficients to keep or discard during the de-noising or compression process. The fixed-form threshold is typically based on a statistical property of the wavelet coefficients and aims to find a balance between noise removal and preserving important signal features.

2. Kernel estimators with fixed and variable bandwidths

The sample of size n drawn from a random variable using the distribution function $f(x)$ is represented by the non-parametric estimate for the dataset $x_1, x_2, \ldots, x_n$. The kernel density estimates at the point x_i are expressed as follows [9]:

$$\hat{f}_h(x_i) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right). \tag{2.1}$$

With kernel functions serving as weights, K represents the kernel probability density function (pdf) centered at each point (x_i). The bandwidth, or smoothing parameter, h, is known as the fixed value, $h > 0$. There are several varieties of kernel density functions [10], with the standard normal distribution being one of the often-employed versions, as demonstrated by the subsequent equation [17]:

$$K(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}. \tag{2.2}$$

When the unknown regression model exhibits consistent behavior throughout the estimate period, the bandwidth parameter h can have a constant value (fixed) over the range of x. The fixed Nadaraya-Watson (FNW) kernel function estimator may be employed in the following ways with a fixed bandwidth to estimate the non-parametric model in equation (1.1):

$$\hat{f}^{FNW}(x_i) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)}, \tag{2.3}$$

where $K(.)$ is the kernel pdf (standard normal distribution), which may be obtained from equation 2.2. There are several methods for selecting the fixed h, including cross-validation and Läuter [17] recommendations, or using optimal methods for selecting the best smoothing parameter, as is the case in this study.

Working with long-tailed and multi-modal distributions is not advised when using the FNW kernel estimator [22]. It is better to choose to change the bandwidth instead. The variable bandwidth estimator $h(x_i)$ for the Variable Nadaraya-Watson (VNW) kernel function may be used as follows to estimate a non-parametric model in Eq. (1.1)

$$\hat{f}^{VNW}(x_i) = \frac{\sum_{i=1}^{n} \frac{y_i}{h(x_i)} K\left(\frac{x-x_i}{h(x_i)}\right)}{\sum_{i=1}^{n} \frac{1}{h(x_i)} K\left(\frac{x-x_i}{h}\right)}. \tag{2.4}$$

Abramson [1] suggested an equation to estimate $h(x_i)$

$$h(x_i) = \frac{h}{\sqrt{f(x_i)}} \tag{2.5}$$

Eq. (2.5) presents the equation for the Abramson estimator, which Läuter [17] refers to as an adaptive kernel function estimator, where $f(x_i)$ is the pdf of x_i that can be calculated using the kernel function estimator. The first stage [16] included using the initial Kernel function estimator with a fixed bandwidth, which is indicated by the local h factor g_i as follows:

$$g_i = \left[\frac{\tilde{f}(x_i)}{g} \right]^{-\alpha}, \tag{2.6}$$

where a is the sensitivity parameter, meaning that $(0 \leq \alpha \leq 1)$, and g is the geometric mean, meaning that $g = 0$. Abramson [1] selected the value $(\alpha = 0.5)$ because it produces strong predicted results. Läuter [17] suggested an adaptable h in the following manner for the second step

$$h(x_i) = hg_i. \tag{2.7}$$

To calculate the h factor in the NW kernel estimator, Demir and Toktamı ş [11] proposed a variation to the (ANW) technique that relies on the arithmetic mean $(\bar{x})$ of $\tilde{f}(x_i)$ rather than g. This is equivalent to

$$\bar{x}_i = \left[\frac{\tilde{f}(x_i)}{\bar{x}} \right]^{-\alpha}. \tag{2.8}$$

Subsequently, the ANW kernel function estimator can be expressed as

$$\hat{f}^{\mathrm{MNW}}(x_i) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\bar{x}_i} K\left(\frac{x-x_i}{h\bar{x}_i}\right)}{\sum_{i=1}^{n} \frac{1}{h\bar{x}_i} K\left(\frac{x-x_i}{h\bar{x}_i}\right)}. \tag{2.9}$$

Using the range (R) of $\tilde{f}(x_i)$, or the difference between the highest and smallest values, as opposed to g or $\bar{x}$ to calculate the bandwidth parameter in the NW kernel estimator, Aljuhani and Al-Turk [7] presented a modification for the Nadaraya-Watson (MNW) technique, that is,

$$\hat{f}^{\mathrm{RNW}}(x_i) = \frac{\sum_{i=1}^{n} \frac{y_i}{hR_i} K\left(\frac{x-x_i}{hR_i}\right)}{\sum_{i=1}^{n} \frac{1}{hR_i} K\left(\frac{x-x_i}{hR_i}\right)}. \tag{2.10}$$

To increase the predictive power of the NW kernel function estimator, Ali [3] suggested a modification to the Nadaraya-Watson (MNW) approach that bases the bandwidth parameter on the median of $\tilde{f}(x_i)$ rather than g, $\bar{x}$ or R. The estimation of the procedure is as follows:

$$\hat{f}^{\mathrm{MeNW}}(x_i) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\mathrm{Me}_i} K\left(\frac{x-x_i}{h\mathrm{Me}_i}\right)}{\sum_{i=1}^{n} \frac{1}{h\mathrm{Me}_i} K\left(\frac{x-x_i}{h\mathrm{Me}_i}\right)}. \tag{2.11}$$

3. Proposed ANW Kernel function estimator based on Dmey and Coiflet wavelets

In this article, we employ two specific types of wavelets, Dmey and Coiflets, for the smoothing and estimation process. Both wavelets are widely used in signal processing; Dmey wavelets are known for their compact support and smoothness, while Coiflets are preferred for their orthogonality and near symmetry, making them appropriate for certain forms of non-parametric regression analysis.

Using wavelet shrinkage (Dmey and Coiflets), a unique method for de-noising non-parametric regression data is provided in this paper. This method proves effective in handling contaminated data and data with long-tailed and multi-modal distributions. The procedure could be summarized into the following main steps: To discover the best goodness fit of the NW kernel function estimator, the bandwidth is then calculated using the NW method, which relies on the fixed-form threshold methodology, which is regarded as $\tilde{f}(x_i)$ rather than using $G, \bar{x}, R, or Me$. The following key steps might be used to summarize the process:

(1) Use Dmey and Coiflet's wavelets to compute the discrete wavelet transform (DWT) on the data.

(2) Estimating median detail coefficients at the first level (Median Absolute Deviation (MAD)):

$$MAD = median\left[\ |W_{1,0}|,\quad |W_{1,1}|,\quad \cdots\quad |W_{1,(\frac{N}{2})-1}|\ \right], j = 1, 2, \cdots, (\frac{N}{2}) - 1 \tag{3.1}$$

Where $(W_{(1,j)}, j = 1, 2, \cdots, \frac{N}{2} - 1$, are the components that symbolize the initial scale of the DWT.

(3) Use the following equation to estimate the universal threshold level

$$\delta_i = \hat{\sigma}(MAD)\sqrt{2\log(N)}, \tag{3.2}$$

where N represents how many wavelet coefficients there are in a certain level, and the sample size must equal $2^j, j = 1, 2, \cdots, N - 1$ and $\hat{\sigma}(MAD)$ is an estimate of the standard deviation of the noise, which can be found by dividing the $\frac{N}{2}$ wavelet coefficients at the first level of decomposition by 0.6745 and applying the MAD estimator to the result [8].

(4) A smooth de-noising can be achieved by using Donoho's algorithm (soft thresholding). The following expression may be used to express the soft threshold de-noising function

$$W_{(J,k)} = \begin{cases} W_{(J,k)}, & |W_{(J,k)}| \geq \delta \\ 0, & |W_{(J,k)}| < \delta \end{cases} \tag{3.3}$$

where $W_{J,k}$ is the transformation coefficient and δ is the threshold. Soft thresholding has a smaller variance than hard thresholding; hence, the primary focus lies in utilizing soft thresholding for modeling and forecasting in this paper.

(5) Calculate the inverse of the DWT

$$x^* = W^{-1}x \tag{3.4}$$

Donoho's algorithm has some fascinating qualities. Soft thresholding ensures that $|x_i^*| < |x_i|$ holds and that x^* has the same smoothness as x. Soft thresholding might be described as the best estimate.

(6) The NW kernel function estimator for de-noised data x^* and fixed-form threshold δ_i will be written as

$$\hat{f}^{\mathrm{FNW}}(x_i) = \frac{\sum_{i=1}^{n} \frac{y_i}{h\delta_i} K\left(\frac{x^* - x_i}{h\delta_i}\right)}{\sum_{i=1}^{n} \frac{1}{h\delta_i} K\left(\frac{x^* - x_i}{h\delta_i}\right)}. \tag{3.5}$$

4. Dmey and coiflets wavelets

The Dmey wavelet (often referred to as Daubechies' wavelet) is a family of wavelets introduced by Ingrid Daubechies, which is commonly used for multi-resolution analysis in signal processing. The Dmey wavelet has compact support, which means it is localized in both time and frequency, making it an excellent choice for analyzing signals with localized features or sharp transitions. The Dmey or "DMeyer" wavelet, also known as the "Daubechies-Meyer" wavelet, is a specific type of wavelet used in wavelet analysis and signal processing. The Daubechies-Meyer wavelet is a member of the Daubechies wavelet family, which is named after Ingrid Daubechies, a prominent mathematician in the field of wavelets. The Daubechies-Meyer wavelet is designed to have compact support and good time-frequency localization properties [5].

The Dmey wavelet is characterized by its smoothness and is particularly useful for representing signals with smooth components or analyzing signals with gradual changes. Numerous applications have used it, including feature extraction, denoising, and picture and audio compression. The exact mathematical equation and properties of the Dmey wavelet depend on the specific order or length of the wavelet, which establishes the number of disappearing moments and the degree of approximation it can attain. Commonly used lengths for the Dmey wavelet include 2, 4, 6, 8, and higher. In this study, we applied the Dmey wavelet decomposition to the data set to capture both low- and high-frequency information. Using the Dmey wavelet at multiple resolution levels, we can isolate and analyze different data components, which are subsequently used for kernel smoothing in the regression process. The compact support of Dmey wavelets ensures that the local features of the signal are well represented without introducing artifacts, making them ideal for our bandwidth estimation approach.

Coiflets are a family of wavelets introduced by [10] as an extension of the Daubechies wavelets, designed to have both compact support and higher regularity. The Coiflets wavelet is known for its orthogonality and near-symmetry, making it ideal for applications requiring precise, high-quality reconstruction of signals. Compared to Daubechies wavelets, Coiflets offer additional vanishing moments, which provide better smoothness while preserving more detail in the signal decomposition. They derive their name from Christopher A. Coifman, a collaborator of mathematician Stéphane G. Mallat, and former student of Ingrid Daubechies. Coiflets wavelets are designed to strike a favorable balance between time and frequency localization, rendering them valuable in a wide range of applications.

Coiflets wavelets are similar to Daubechies wavelets but have slightly different properties [6]. They have a higher number of vanishing moments, allowing them to represent more complex signal features and capture sharper transitions in the signal. Coiflet wavelets also have a smoother scaling function compared to other wavelet families, which makes them suitable for analyzing signals with a high degree of smoothness. The Coiflets family includes different wavelets with varying lengths or orders, such as Coiflet1, Coiflet2, Coiflet3, and so on. Each wavelet in the Coiflets family is associated with a specific scaling function and wavelet function, which determine the characteristics of the wavelet transformation. It is important to remember that the wavelet selection is based on the needs of the application and the properties of the signal being examined. Coiflet wavelets, with their specific properties, provide an alternative to other wavelet families, and can be beneficial in scenarios where their characteristics align with the desired analysis goals.

5. Evaluation criteria

The MSE will be used as an evaluation criterion to evaluate the classical and suggested NW kernel estimators. Its foundation lies in calculating the difference between the observed values y_i and anticipated $\hat{f}(x_i)$ using the suggested and traditional NW kernel estimators. The estimator with the lowest MSE value is the best. The following are some possible applications for the MSE in non-parametric regression:

$$MSE = \frac{1}{tr(I-V)} \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2, \tag{5.1}$$

where I is the identity matrix and $V = (I - \alpha K)^{-1}$. In parametric regression, the conventional procedure is to divide the sum of squared errors by the degrees of freedom. The average MSE is calculated by repeating (m) times as [2]

$$AMSE = \frac{\sum_{j=1}^{m} MSE_j}{m} \tag{5.2}$$

6. Simulation results

Using MATLAB (R2021a) language software created especially for this purpose, the performance of the estimators in the proposed method was compared using a simulation analysis with the classical Nadaraya-Watson estimators. To demonstrate the procedure for creating upper and lower locally weighted running line smoothers, synthetic data was created throughout the simulation. A sine wave was given noise, as shown in the following non-linear regression function, to get this data

$$y_i = \sin(x_i) + 0.75 * \varepsilon_i \sim N(0,1) \quad ; i = 1, 2, \dots, n. \tag{6.1}$$

in which x_i was selected using a MATLAB "linspace" algorithm based on a uniform distribution over the interval $[0, 4 \times \pi]$. Thus, sample sizes of 32, 64, 128, and 256 were generated. The preset h values were 1, 0.75, and 0.5. Using the usual normal kernel function, the kernel estimates for the variables NW-geometric, ANW-mean, ANW-range, ANW-median and suggested approaches (ANW-Dmey and Coiflets) were computed. The average MSE for each approach, derived from 1000 repeated simulations, is displayed in Table 1. The generated data and the estimates of the conventional and recommended nonparametric regression approaches for samples of sizes of 32, 64, 128 and 256 are shown in Figures 1, 2, 3, and 4 (at $h = 0.5$).

Table 1. The mean MSE for the suggested Dmey and Coiflets and conventional methods

n	H	Geometric	Mean	Range	Median	Dmey	Coiflets
32	0.50	0.4988	0.4491	0.3141	0.4591	0.1870	0.1848
64	0.50	0.6573	0.6301	0.5524	0.6304	0.2077	0.2051
128	0.50	0.7416	0.7267	0.6874	0.7241	0.2858	0.2843
256	0.50	0.7841	0.7750	0.7542	0.7720	0.3428	0.3426
32	0.75	0.6284	0.5785	0.4602	0.5855	0.2261	0.2225
64	0.75	0.7434	0.7123	0.6425	0.7114	0.2832	0.2790
128	0.75	0.8024	0.7803	0.7375	0.7763	0.3483	0.3459
256	0.75	0.8314	0.8136	0.7837	0.8085	0.3876	0.3870
32	1.00	0.7362	0.6746	0.5460	0.6842	0.2949	0.2912
64	1.00	0.8279	0.7834	0.6969	0.7847	0.3508	0.3459
128	1.00	0.8726	0.8361	0.7719	0.8332	0.4015	0.3987
256	1.00	0.8950	0.8622	0.8087	0.8570	0.4290	0.4283

Table 1 presents the simulation results that show how adaptive nonparametric techniques of the NW kernel (for geometric, arithmetic mean, range, and median) may be used for the estimation of nonlinear regression models. These are flexible approaches that have been shown to produce precise prediction outcomes in the past. A novel NW kernel function estimator has been proposed as an enhancement of the adaptive NW kernel estimator. This estimator is based on enhancing the predictive power of the ANW kernel estimator by utilizing the universal threshold level to increase the local factor of the bandwidth instead of the arithmetic mean, range, geometric, and median when calculating the bandwidths, and Dmey and Coiflets wavelet to de-noise the data. In this simulation study, various sample sizes and initial bandwidth values were utilized to compare the estimators of the proposed method with those of the classical methods, based on the average MSE criterion.

The results revealed that the proposed Coiflets method outperformed the Dmey wavelet method. In practice, the mean MSE increases with sample size due to the application of nonparametric regression and the data's tendency to resemble a normal distribution as the sample size increases, or in certain cases, increasing the sample size could result in overfitting, particularly if the estimator is too complex for the data, leading to a rise in MSE.

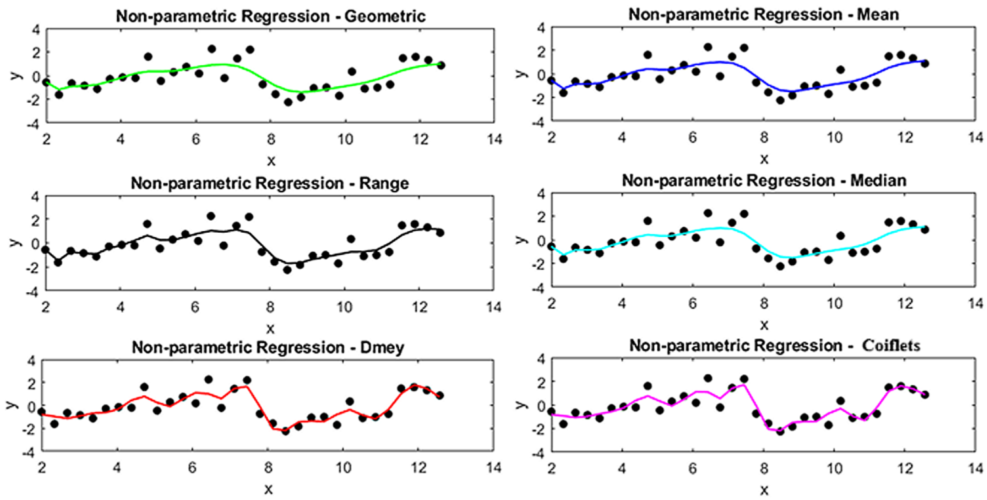

Figure 1. Traditional and suggested techniques at $h = 0.5, n = 32$

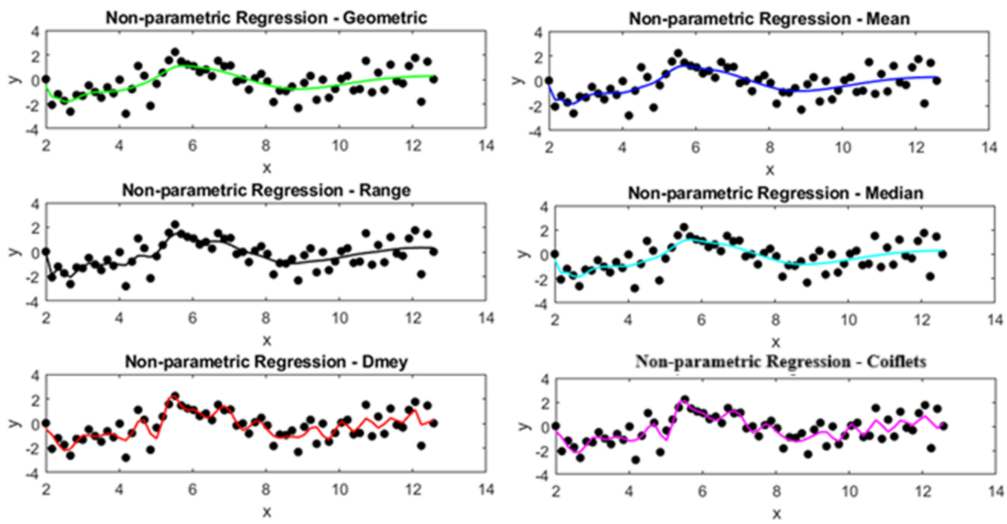

Figure 2. Traditional and suggested techniques at $h = 0.5, n = 64$

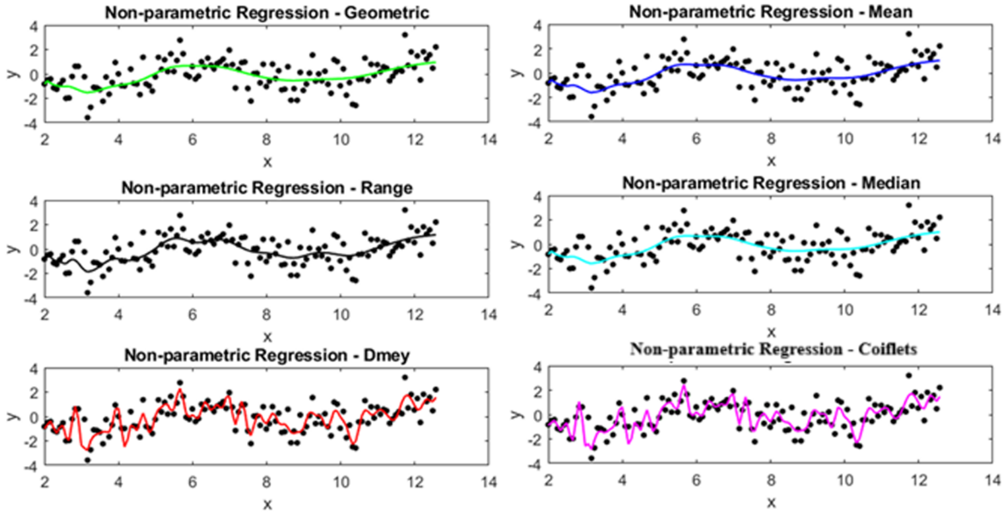

Figure 3. Traditional and suggested techniques at $h = 0.5, n = 128$

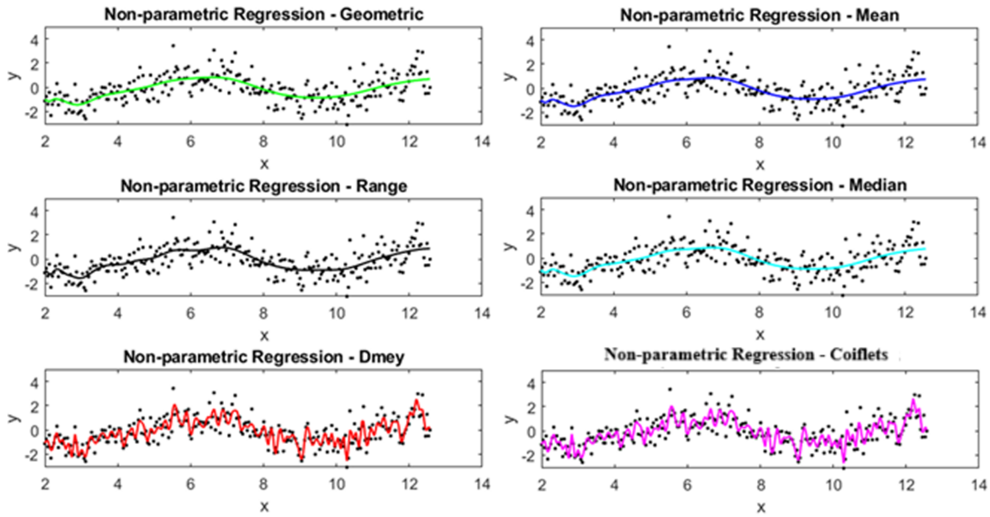

Figure 4. Traditional and suggested techniques at $h = 0.5, n = 256$

7. Application

Data on the daylight outbursts of Old Faithful Geyser in Yellowstone National Park during August 1 to 4, 1978 (for the first 32 observations from 52) [25]. The variables are y = interval till the next eruption and x = duration of an eruption. A non-parametric regression may be used to effectively predict variable y by applying variable x. Due to its inadequate circumstances, simple regression cannot be used. Using non-parametric regression techniques is one way to solve the problem; as such, Figures 5, 6, and 7 (at $h = 0.2, 0.3, and 0.5$) show the classical and suggested ways, respectively.

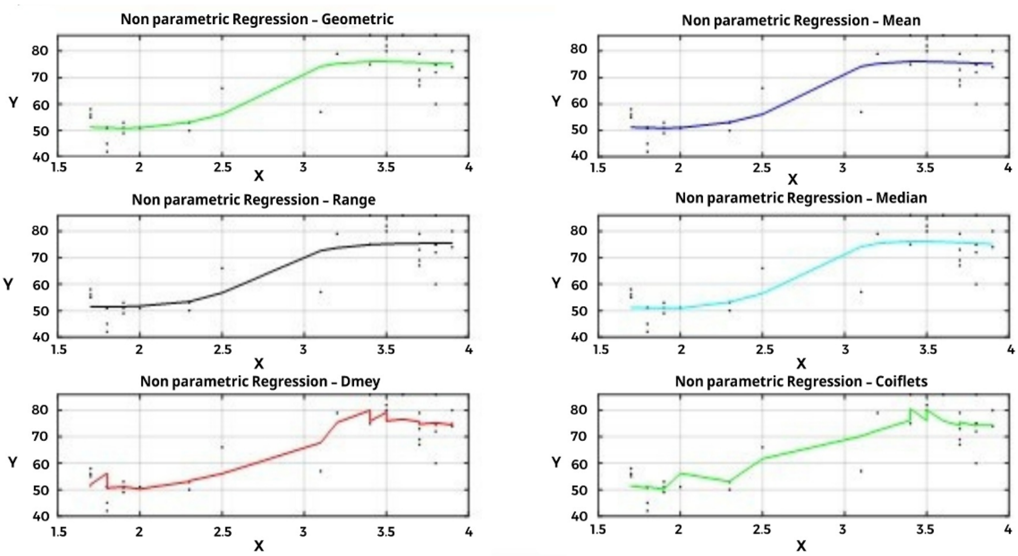

Figure 5. Traditional and suggested methods at $h = 0.2$ for real data

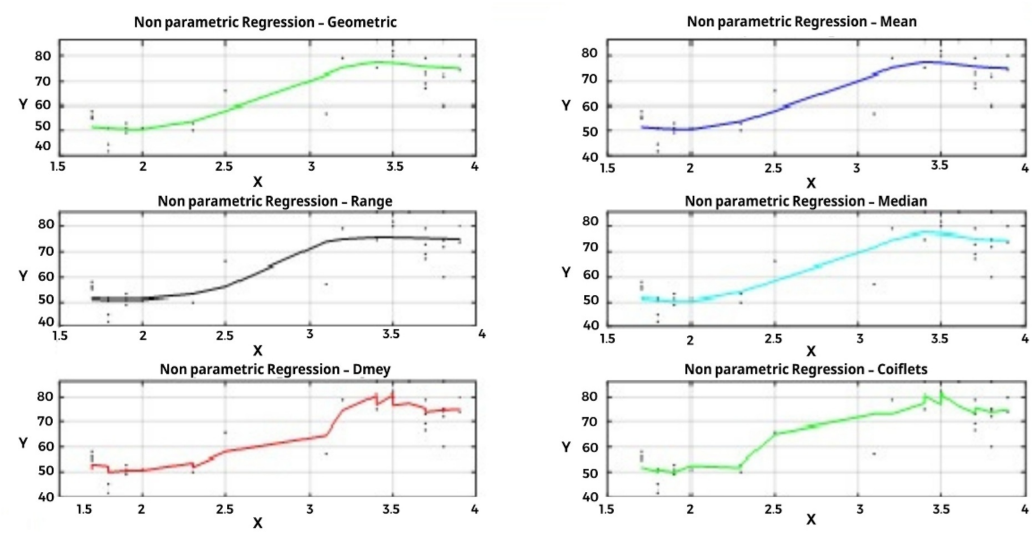

Figure 6. Traditional and suggested methods at $h = 0.3$ for real data

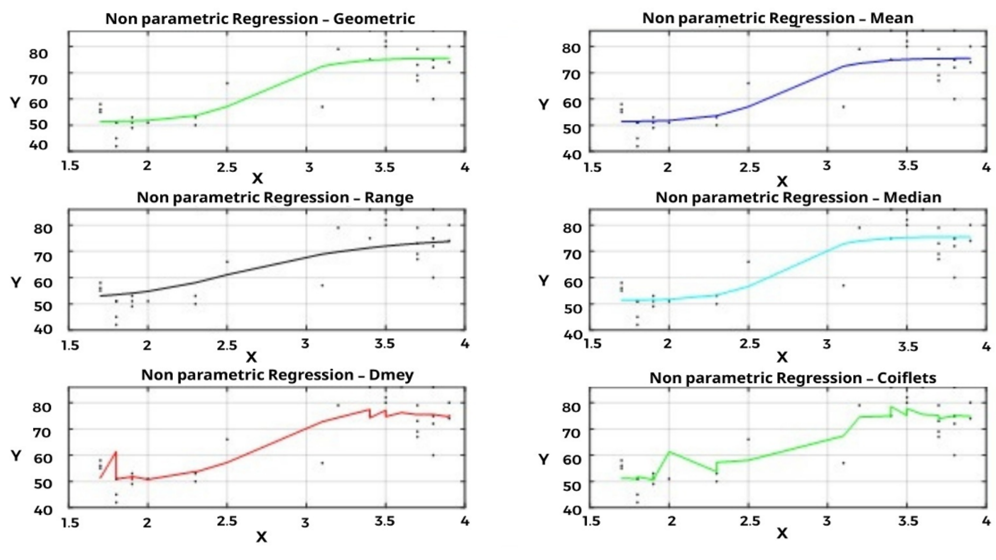

Figure 7. Traditional and suggested methods at $h = 0.5$ for real data

The suggested new NW kernel function estimator, which incorporates an adjustment to the ANW kernel estimator, shows better predictive ability than the ANW kernel estimator based on the findings of the real data shown in Table 2. This improvement is made possible by utilizing the Dmey and Coiflets wavelets for data de-noising and increasing the local factor of the bandwidth while computing the bandwidths using the universal threshold level. A detailed study was carried out using different values of the starting bandwidth, and the estimators of the suggested method were contrasted with estimators of classical techniques using MSE criteria. The results unequivocally show that the suggested strategy is more effective and efficient than conventional methods since it produces minimal MSE values.

Table 2. MSE of the traditional and suggested methods with actual data

h	Geometric	Mean	Range	Median	Dmey	Coiflets
0.2	39.2524	39.1132	44.0098	37.9888	34.3570	34.2965
0.3	43.8581	43.8068	44.0522	43.3072	41.9333	34.9105
0.5	43.9936	44.0010	52.8263	44.1126	42.8884	39.0000
1.0	66.9735	66.7205	134.703	64.1597	62.8748	58.0701
1.5	110.093	110.013	168.476	108.545	83.8393	74.0854
2.0	136.517	136.497	176.323	135.893	97.7812	89.0199

8. Conclusion

The proposed method estimator, which uses the universal threshold level with Coiflets and Dmey wavelets, outperformed all classical methods for both simulated and actual data, as assessed by the MSE criterion. Among the estimators of the proposed method, the one that used the universal threshold level with the Coiflets wavelet outperformed the Dmey wavelet method. Reducing the starting bandwidth values improved the predictive performance of the estimators. More accurate predictions were obtained by optimizing the estimators using smaller sample sizes.

Implement the proposed method for estimating bandwidth in the NW kernel of non-parametric regression estimators, using the universal threshold level with Coiflets and Dmey wavelets for data de-noising instead of relying on variable bandwidths such as geometric, arithmetic mean, range, and median. Future research should be done to estimate the bandwidth using alternative wavelets or varying degrees of universal threshold. Conduct further studies to explore the effectiveness of using different levels of universal threshold or other wavelet functions for bandwidth estimation to identify the most suitable approach for specific datasets and applications.

It is recommended to consider using the Maximal Overlap Discrete Wavelet Transformation (MODWT) instead of DWT for bandwidth estimation. This method may offer advantages when dealing with sample sizes that are not always equal to (2^j). Finally, investigate the estimation of bandwidth using different thresholding methods such as Minimax, Stein's Unbiased Risk Estimate (SURE), or Bayesian approaches in combination with wavelets. These methods may provide useful information for improving the accuracy of the bandwidth estimate.

Acknowledgements

Sincere thanks are extended by the authors to the reviewers for their intelligent recommendations and helpful criticism, which have significantly increased the caliber of this work. We also thank the editorial staff of the Hacettepe Journal of Mathematics and Statistics for their invaluable help and advice during the publishing process.

Author contributions. All the co-authors have contributed equally in all aspects of the preparation of this submission.

Conflict of interest statement. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding. The authors declare that no funding or financial support was received during the preparation of this manuscript.

Data availability. The dataset used in this study was obtained from the book *Applied Linear Regression* by S. Weisberg (Wiley, New York, 1985).

References

[1] I. Abramson, *On bandwidth variation in kernel estimates—a square root law*, Ann. Stat. **10**:1217–1223, 1982.

[2] T. H. Ali, *Using proposed nonparametric regression models for clustered data (a simulation study)*, ZANCO J. Pure Appl. Sci. **29**:78–87, 2017.

[3] T. H. Ali, *Modification of the adaptive Nadaraya-Watson kernel method for nonparametric regression (simulation study)*, Commun. Stat. Simul. Comput. **51**:391–403, 2022.

[4] T. H. Ali, H. A. A.-M. Hayawi, and D. Shaker Botani, *Estimation of the bandwidth parameter in Nadaraya-Watson kernel non-parametric regression based on universal threshold level*, Commun. Stat. Simul. Comput. **52**:1476–1489, 2023.

[5] T. H. Ali and J. R. Qadir, *Using wavelet shrinkage in the Cox proportional hazards regression model (simulation study)*, Iraq J. Stat. Sci. **19**:17–29, 2022.

[6] T. H. Ali and D. M. Saleh, *Comparison between wavelet Bayesian and Bayesian estimators to remedy contamination in linear regression model*, PalArch J. Egypt. Egyptol. **18**, 2021.

[7] K. H. Aljuhani and L. I. A. Turk, *Modification of the adaptive Nadaraya-Watson kernel regression estimator*, Sci. Res. Essays **9**:966–971, 2014.

[8] I. L. Cascio, *Wavelet analysis and denoising: New tools for economists*, 2007.

[9] A. Christmann and I. Steinwart, *Consistency and robustness of kernel-based regression in convex risk minimization*, Bernoulli **13**:799–819, 2007.

[10] R.R. Coifman and M.V. Wickerhauser, *Entropy-based algorithms for best basis selection*, IEEE Trans. Inf. Theory **38**(2):713–718, 1992.

[11] S. Demir and Toktamı ş, *On the adaptive Nadaraya-Watson kernel regression estimators*, Hacet. J. Math. Stat. **39**:429–437, 2010.

[12] R. Eubank, *Spline Smoothing and Nonparametric Regression*, Dekker, 1988.

[13] J. H. Friedman and W. Stuetzle, *Projection pursuit regression*, J. Am. Stat. Assoc. **76**:817–823, 1981.

[14] Y. A. Hassan and M. Y. Hmood, *Estimation of return stock rate by using wavelet and kernel smoothers*, Period. Eng. Nat. Sci. **8**(2):602–612, 2020.

[15] W. Härdle and G. Kelly, *Non-parametric kernel regression estimation optimal choice of bandwidth*, Stat. **18**:21–35, 1987.

[16] D. Li and R. Li, *Local composite quantile regression smoothing for Harris recurrent Markov processes*, J. Econom. **194**:44–56, 2016.

[17] H. Läuter, *Silverman, B. W.: Density estimation for statistics and data analysis. Chapman & Hall, London – New York, 1986*, Biometr. J. **30**:876–877, 1988.

[18] M. Y. Mustafa and Z. Y. Algamal, *Smoothing parameter selection in kernel nonparametric regression using bat optimization algorithm*, J. Phys.: Conf. Ser. **1897**, 2021.

[19] E. A. Nadaraya, *On estimating regression*, Theory Probab. Appl. **9**:141–142, 1964.

[20] D. H. Rashid, M. Y. Hmood, and S. K. Hamza, *Nadaraya-Watson estimator a smoothing technique for estimating regression function*, J. Econ. Adm. Sci. **18**(65):283, 2012.
[21] D. W. Scott and G. R. Terrell, *Biased and unbiased cross-validation in density estimation*, J. Am. Stat. Assoc. **82**:1131–1146, 1987.
[22] S. Shahzadi, U. Shahzad, and N. Koyuncu, *On the adaptive Nadaraya-Watson kernel estimator for the discontinuity in the presence of jump size*, SDU J. Nat. Appl. Sci. **22**:511–520, 2018.
[23] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, 1995.
[24] G. S. Watson, *Smooth regression analysis*, Sankhyā Ser. A **26**:359–372, 1964.
[25] S. Weisberg, *Applied Linear Regression*, Wiley, New York, 1985.