# Bandwidth Selection in Multivariate Nadaraya-Watson Estimator based on Meta-Heuristic Optimization Algorithms: A Simulation Study

**Marwah Yahya Mustafa**
Department of Statistics and Informatics, University of Mosul, Mosul, Iraq
Email: marwa.yehea1987@gmail.com

**Zakariya Yahya Algamal (Corresponding Author)**
Department of Statistics and Informatics, University of Mosul, Mosul, Iraq
E-mail: zakariya.algamal@uomosul.edu.iq
ORCID: 0000-0002-0229-7958

***Article Info***
***Page Number: 4877 - 4887***
***Publication Issue:***
***Vol 71 No. 4 (2022)***

***Article History***
***Article Received***: *25 March 2022*
***Revised***: *30 April 2022*
***Accepted***: *15 June 2022*
***Publication***: *19 August 2022*

**Abstract**

The smoothing parameter has a complete bearing on curve estimation in the context of kernel nonparametric regression.Meta-heuristic algorithm implementation has grown in popularity among researchers.In this study, a multivariate Nadaraya-Watson kernel nonparametric regression bandwidth matrix selection approach based on a pigeon optimization algorithm is given.The suggested approach will effectively assist in identifying the appropriate bandwidth matrix with a strong forecast.The proposed approach is contrasted with two well-known approaches.The thorough demonstration of the suggested method's superiority in terms of prediction ability is provided by the experimental results.

**Keywords**: kernel estimator; smoothing matrix, Nadaraya-Watson estimator, meta-heuristic algorithm.

## 1.Introduction

Nonparametric kernel estimation method is the most popular nonparametric estimationmethod(Härdle, 1990). Nonparametric approaches to the smoothing of noisy regression data sets have been demonstrated to be effective for many applications. “For detailed expositions see Eubank (1988)Eubank , Müller (1988)and Härdle (1990). An important component of all nonparametric regression estimators is the choice of the smoothing parameter . The nonparametric regression model, often estimated by estimators of the Nadaraya–Watson type (Nadaraya (1964), Watson (1964)) . The estimation resulte, however, depend crucially on the choice of bandwidth . In a univariate case these estimates depend on a bandwidth (h), which is a smoothing parameter controlling smoothness of an estimated curve and a kernel (k) which is considered as a weight function .

The choice of the smoothing parameter is a crucial problem in the kernel regression. The literature on bandwidth selection is quite extensive, such as (Ali, 2019; Chen, 2015; C.-K. Chu & Marron, 1991; C. Chu, 1995; Dobrovidov & Ruds’ko, 2010; Feng & Heiler, 2009; Francisco-Fernández & Vilar-Fernández, 2005; Gao & Gijbels, 2012; Kauermann & Opsomer, 2004; Koláček & Horová, 2017; Lee & Solo, 1999; Leungi, Marriott, & Wu, 1993; Mustafa & Algamal, 2021; Nychka, 1991; Opsomer & Miller, 2007; Rice, 1984; Schucany, 1995; Zhang, Chan, Ho, & Ho, 2008; Zhou & Huang, 2018; Żychaluk, 2014). For the multivariate case (d-dimemsional), these estimates depend on a bandwidth matrix $\mathbf{H}$ and $K(\cdot)$ is a multivariate kernel function”.

In this paper, meta-heuristic optimization algorithms method proposed to choose smoothing parameter in multivariate Nadaraya-Watson kernel nonparametric regression. The proposed method will efficiently help to findsuperiority of the proposed method in different simulated examples and a real data application is proved. the best smoothing parameter with a high prediction.

## 2. Multivariate Nadaraya-Watson estimator (MNW)

Consider the multivariate nonparametric regression model MNR defined as (Schafer & Wasserman, 2013)

$$\mathbf{y_i} = \mathbf{m}(\mathbf{X}) + \boldsymbol{\varepsilon}_\mathbf{i} \qquad i = 1, 2, \ldots, n \tag{1}$$

where $\mathbf{y}$ is a scalar dependent variable, $\mathbf{X}=(\mathbf{x_{1i}},\mathbf{x_{2i}},\ldots,\mathbf{x_{di}})$ is a matrix of independent d-variables, $\boldsymbol{\varepsilon}_\mathbf{i}$ is independent of $\mathbf{X}$, $E(\boldsymbol{\varepsilon})=0$ and $\operatorname{var}(\boldsymbol{\varepsilon})=\sigma^2$ and $m(\mathbf{x})$ is the function of unknown regression estimated by Multivariate Nadaraya-Watson estimator (Multivariate local constant) is defined as the following :

$$\hat{m}(\mathbf{x})_{MNW}=\frac{\sum_{i=1}^{n}\mathbf{K}\left(\mathbf{H}^{-1}(\mathbf{x_i}-\mathbf{x})\right)\quad\mathbf{y_i}}{\sum_{i=1}^{n}\mathbf{K}\left(\mathbf{H}^{-1}(\mathbf{x_i}-\mathbf{x})\right)} \tag{2}$$

Where $K(\square)$ is a multivariate kernel function and $\mathbf{H}$ is a symmetric positive definite matrix called the bandwidth matrix,where $\mathbf{h}=[h_1\ h_2\cdots h_d]'$, $\mathbf{H}=\mathbf{h}_{1\times d}\mathbf{I}_{d\times d}$, then

$$\mathbf{H}=\begin{bmatrix} h_1 & 0\cdots 0 \\ & h_2\cdots 0 \\ & \ddots\ \vdots \\ & & h_d \end{bmatrix}_{d\times d} \tag{3}$$

The multivariate Nadaraya-Watson estimator (MNW) depend crucially on the optimal bandwidth matrix choice .There are many methods to choice bandwidth matrix $\mathbf{H}$, Cross validation (CV), General cross validation(GCV), Kullback-Leibler cross validation (KLCV) and Least –Squares cross validation (LSCV), we measure the performance of $\hat{m}(\mathbf{x})$ by mean integrated squared error (MISE).

### 3. The proposed method

The efficiency of MNW kernel estimator largely depends on an appropriately choosing the smoothing parameter matrix , $\mathbf{H}$. “As a result, it is of crucial importance selecting a suitable value of the $\mathbf{H}$ . In literature, the most widely used method for selecting $\mathbf{H}$ is the cross-validation (CV) and General cross validation(GCV) .

$$CV(\mathbf{H})=\frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{y_i}-\hat{m}_{-1}(\mathbf{X_i};\mathbf{k})\right]^2 \tag{4}$$

$$\hat{\mathbf{H}}=\arg\min_{H\in H}\mathbf{H} \tag{5}$$

$$GCV_{\mathbf{H}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\hat{m}\left(\mathbf{x_i}\right)-\mathbf{y_i}}{1-tr\left(S_{\mathbf{H}}\right)/n}\right\}^2 \tag{6}$$

$$\hat{\mathbf{H}}_{Gcv} = \underset{H \in H}{argmin}\, \mathrm{GCV}_H \tag{7}$$

Where $\hat{m}_{-1}(\mathbf{X_i;k})$ is multivariate Nadaraya-Watson estimator (MNW) and n represent the size of sample. In this paper, a meta-heuristic optimization algorithms method is proposed to determine the smoothing parameter matrix in the MNW kernel estimator. The proposed method will efficiently help to find the best value with high prediction performance. The parameter configurations for our proposed method are presented as follows.

Nature has been an inspiration for the introduction of many meta-heuristic algorithms. Swarm intelligence is an important tool for solving many complex problems in scientific research (Algamal, 2018; Kahya, Altamir, & Algamal, 2020). Swarm intelligence algorithms have been widely studied and successfully applied to a variety of complex optimization problems.

The pigeon optimization algorithm (POA), which was proposed by Duan and Qiao (2014), has certain outstanding merits, such as a simple computational process, simple implementation, and easy understanding with only a few parameters for tuning. Due to its good properties, POA has become a useful tool for many real-world problems (Fu, Chan, Niu, Chung, & Qu, 2019; Qiu & Duan, 2020; Sushnigdha & Joshi, 2018; Yan, Qu, Zhu, Qiao, & Suganthan, 2019; Yang, Duan, Fan, & Deng, 2018; Zhong, Wang, Lin, & Zhang, 2019). The POA simulates the homing behavior of pigeons.

The POA mainly consists of two operators: the map and compass operator and the landmark operator. In the map and compass operator, pigeons sense the geomagnetic field to shape the map for homing. Suppose that the search space is N-dimensional, and then the i-th pigeon of the swarm can be represented by a N-dimensional vector $X_i = (X_{i,1}, X_{i,2}, ..., X_{i,\mathrm{N}})$. The velocity of this pigeon, which represents the position change of this pigeon, can be represented by another N-dimensional vector $V_i = (V_{i,1}, V_{i,2}, ..., V_{i,\mathrm{N}})$. The best previously visited position of the i-th pigeon is denoted as $P_i = (P_{i,1}, P_{i,2}, ..., P_{i,\mathrm{N}})$.The global best position of the swarm is $g = \left(g_1, g_2, \ldots, g_N\right)$. Each pigeon is flying according to the following two equations:

$$V_i\left(t+1\right) = V_i\left(t\right) \times e^{-Rt} + rand \times \left(X_g - X_i\left(t\right)\right) \tag{8}$$

$$X_i(t+1) = X_i(t) + V_i(t+1), \tag{9}$$

where $R$ is a map and compass factor, while $rand$ is a uniform random number in the range $[0, 1]$, $X_g$ is the global best solution, $X_i(t)$ denotes the current position of a pigeon at instance $t$, and $V_i(t)$ denotes the current velocity of a pigeon at iteration $t$.

In landmark operator, all the pigeons are ranked according to their fitness value. In each generation, the number of pigeons is updated by Eq. (9), where only half number of pigeons is considered to calculate the desired position of the centered pigeon, while all other pigeons adjust their destination by following the desirable destination position.

$$N_p(t+1) = \frac{N_p(t)}{2}, \tag{10}$$

where $N_p$ is the number of pigeons in the current iteration $t$.

The position of the desired destination is calculated by Eq. (8), while all other pigeons update their position toward this position by Eq. (9) (Duan & Qiao, 2014).

$$X_c(t+1) = \frac{\sum X_i(t+1) \times \text{Fitness}(X_i(t+1))}{N_p \sum \text{Fitness}(X_i(t+1))} \tag{11}$$

$$X_i(t+1) = X_i(t) + rand \times \left(X_c(t+1) \times X_i(t)\right), \tag{12}$$

where $X_c$ is the position of the centered pigeon (desired destination).

The efficiency of MNW kernel estimator largely depends on an appropriately choosing the smoothing parameter matrix, $H$. As a result, it is of crucial importance selecting a suitable value of the $h_i$. In this paper, a POA is proposed to determine the smoothing parameter in the MNW kernel estimator. The proposed method will efficiently help to find the best value with high prediction performance". The parameter configurations for our proposed method are presented as follows.

(1) The number of pigeons, is set to 30 and the number of iterations is $t_{max} = 250$.

(2) The positions of each pigeon are randomly determined. The position of a pigeon represents the smoothing parameter matrix, $H$ . The initial positions of the pigeon s are generated from a uniform distribution within the range [0,15].

(3) The fitness function is defined as

$$\text{fitness} = \min \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{13}$$

(4) The positions are updated using Eq. (12).

(5) Steps 3 and 4 are repeated until a $t_{\max}$ is reached.

## 4. Simulation results

To test how well the proposed method performs for different possible mean functions the following study design was followed. “The comparisons with different used methods, CV, GCV, are also conducted. “Three sizes of samples are taken as: $n = 50, 100, 250$ In addition, the type of kernel is setting as Epanechnikov kernel typ.

**Case 1:** In the case, we use the regression function(Koláček & Horová, 2017)

$$y_i = (x_1 - 0.5)^3 + (x_2 - 0.5) + \varepsilon_i \quad \sim N(0, 0.02) \tag{14}$$

The independent variables, are generated from uniform distribution with the range 0 and 1.

**Case 2:** In the case, we use the regression function(Shang, Zhang, & Shang, 2014)

$$y_i = \sin\left(2\pi x_{1,i}\right) + 4\left(1 - x_{2,i}\right)\left(1 + x_{2,i}\right) + \frac{2x_{3,i}}{1 + 0.8x_{3,i}^2} + \varepsilon_i \sim N(0, 0.09^2) \tag{15}$$

The independent variables, are generated from uniform distribution with the range 0 and 1.

**Case 3:** In the case, we use the regression function(Lijian & Rolf, 1999)

$$y_i = \left\{\left(x_1 - 0.5\right)^2 + x_2^2\right\} \sin\left(2\pi x_3\right) + \varepsilon_i \quad \sim N(0, 0.25) \tag{16}$$

The independent variables, are generated from uniform distribution with the range 0 and 1.

**Case 4:** In the case, we use the regression function(Lijian & Rolf, 1999)

$$y_i = \sin\left\{\pi\left(x_1 + 0.5x_2 + 2x_3 + 0.5x_4\right)\right\} + \varepsilon_i \quad \sim N(0, 0.25) \tag{17}$$

The independent variables, are generated from uniform distribution with the range 0 and 1.

**Case 5:** In the case, we use the regression function(Goutte, Larsen, & technology, 2000)

$$y_i = 10\sin(\pi x_1 x_2) + 20\left(x_3 - \frac{1}{2}\right)^2 + 10x_4 + 5x_5 + \varepsilon_i \quad \sim N(0,1) \qquad (18)$$

The independent variables, are generated from uniform distribution with the range 0 and 1.

The generated data is repeated 500 times and the averaged integrated mean squared error (IMSE) is calculated. The results of the used methods are summarized in Tables 1-5.

As shown in Tables 1-5, on the five cases, according to the results of the POA and other methods, it can be clearly seen that the results achieved is better than the CV and the GCV. For example, for case 2 and n=250, the reduction in MSE for POA compared to CV and GCV was 57.25% and 55.86%, respectively". Further, regardless of the value of n, the proposed methods, POA, provides considerably better results compared to the other methods, in terms of MSE.

Table 1: AVERAGE MSE values for the used methods of case 1

| | CV | GCV | POA |
|---|---|---|---|
| n=50 | 0.035 ± 0.008 | 0.029 ± 0.011 | **0.004094 ± 0.001856** |
| n=100 | 0.036 ± 0.007 | 0.034 ± 0.007 | **0.007698 ± 0.002252** |
| n=250 | 0.037 ± 0.003 | 0.036 ± 0.004 | **0.015895 ± 0.001955** |

Table 2: MSE values for the used methods of case 2

| | CV | GCV | POA |
|---|---|---|---|
| n=50 | 2.782 ± 0.813 | 1.821 ± 0.905 | **0.022582 ± 0.046602** |
| n=100 | 2.869 ± 0.627 | 2.485 ± 0.710 | **0.058222 ± 0.046109** |
| n=250 | 2.932 ± 0.259 | 2.807 ± 0.276 | **0.101367 ± 0.042885** |

Table 3: MSE values for the used methods of case 3

| | CV | GCV | POA |
|---|---|---|---|
| n=50 | 0.207 ± 0.050 | 0.145 ± 0.072 | **0.001 ± 0.003** |
| n=100 | 0.217 ± 0.044 | 0.187 ± 0.055 | **0.004 ± 0.003** |
| n=250 | 0.224 ± 0.019 | 0.213± 0.023 | **0.007 ± 0.003** |

Table 4: MSE values for the used methods of case 4

| | CV | GCV | POA |
|---|---|---|---|
| n=50 | 0.411 ± 0.159 | 0.061 ± 0.065 | **$1.14\times10^{-7}$ ± $5.71\times10^{-7}$** |
| n=100 | 0.362 ± 0.138 | 0.111 ± 0.039 | **$9.57\times10^{-6}$ ± $2.70\times10^{-5}$** |
| n=250 | 0.235 ± 0.053 | 0.127 ± 0.033 | **$4.03\times10^{-5}$ ± 0.000132** |

Table 5: MSE values for the used methods of case 5

| | CV | GCV | POA |
|---|---|---|---|
| n=50 | 8.457 ± 3.644 | 1.465 ± 1.090 | **$2.02\times10^{-6}$ ± $1.01\times10^{-5}$** |
| n=100 | 5.695 ± 1.469 | 1.898 ± 1.087 | **0.000209 ± 0.000646** |
| n=250 | 4.408 ± 0.690 | 1.873 ± 0.537 | **0.000746 ± 0.002415** |

## 5. Conclusion

Meta-heuristic algorithm implementation has grown in popularity among researchers.The issue of choosing a smoothing parameter for multivariate Nadaraya-Watson kernel nonparametricregression is discussed in this study.It was suggested to use a pigeon optimization approach to select the smoothing parameter matrix.The simulation results showed that the POA approach, when compared to other competing methods, was superior in terms of IMSE.

## References

1. Algamal, Z. Y. (2018). A new method for choosing the biasing parameter in ridge estimator for generalized linear model. *Chemometrics and Intelligent Laboratory Systems, 183*, 96-101. doi:10.1016/j.chemolab.2018.10.014
2. Ali, T. H. (2019). Modification of the adaptive Nadaraya-Watson kernel method for nonparametric regression (simulation study). *Communications in Statistics - Simulation and Computation*, 1-13. doi:10.1080/03610918.2019.1652319
3. Chen, S. (2015). Optimal Bandwidth Selection for Kernel Density Functionals Estimation. *Journal of Probability and Statistics, 2015*, 1-21. doi:10.1155/2015/242683
4. Chu, C.-K., & Marron, J. S. (1991). Comparison of two bandwidth selectors with dependent errors. *The Annals of Statistics, 19*(4), 1906-1918.

5. Chu, C. (1995). Bandwidth selection in nonparametric regression with general errors. *Journal of statistical planning and inference, 44*(3), 265-275.
6. Dobrovidov, A. V., & Ruds'ko, I. M. (2010). Bandwidth selection in nonparametric estimator of density derivative by smoothed cross-validation method. *Automation and Remote Control, 71*(2), 209-224. doi:10.1134/s0005117910020050
7. Duan, H., & Qiao, P. (2014). Pigeon-inspired optimization: a new swarm intelligence optimizer for air robot path planning. *International Journal of Intelligent Computing and Cybernetics, 7*(1), 24-37. doi:10.1108/ijicc-02-2014-0005
8. Eubank, R. L. (1988). *Spline smoothing and nonparametric regression* (Vol. 90): M. Dekker New York.
9. Feng, Y., & Heiler, S. (2009). A simple bootstrap bandwidth selector for local polynomial fitting. *Journal of Statistical Computation and Simulation, 79*(12), 1425-1439. doi:10.1080/00949650802352019
10. Francisco-Fernández, M., & Vilar-Fernández, J. M. (2005). Bandwidth selection for the local polynomial estimator under dependence: a simulation study. *Computational Statistics, 20*(4), 539-558.
11. Fu, X., Chan, F. T. S., Niu, B., Chung, N. S. H., & Qu, T. (2019). A multi-objective pigeon inspired optimization algorithm for fuzzy production scheduling problem considering mould maintenance. *Science China Information Sciences, 62*(7). doi:10.1007/s11432-018-9693-2
12. Gao, J., & Gijbels, I. (2012). Bandwidth Selection in Nonparametric Kernel Testing. *Journal of the American Statistical Association, 103*(484), 1584-1594. doi:10.1198/016214508000000968
13. Goutte, C., Larsen, J., & technology, v. (2000). Adaptive metric kernel regression. *Journal of VLSI signal processing systems for signal, image, 26*(1-2), 155-167.
14. Härdle, W. (1990). *Applied nonparametric regression*. Cambridge: Cambridge university press.
15. Kahya, M. A., Altamir, S. A., & Algamal, Z. Y. (2020). Improving firefly algorithm-based logistic regression for feature selection. *Journal of Interdisciplinary Mathematics, 22*(8), 1577-1581. doi:10.1080/09720502.2019.1706861
16. Kauermann, G., & Opsomer, J. D. (2004). Generalized Cross-Validation for Bandwidth Selection of Backfitting Estimates in Generalized Additive Models. *Journal of Computational and Graphical Statistics, 13*(1), 66-89. doi:10.1198/1061860043056

17. Koláček, J., & Horová, I. (2017). Bandwidth matrix selectors for kernel regression. *Computational Statistics, 32*(3), 1027-1046. doi:10.1007/s00180-017-0709-3
18. Lee, T. C., & Solo, V. (1999). Bandwidth selection for local linear regression: a simulation study. *Computational Statistics, 14*(4), 515-532.
19. Leungi, D. H. Y., Marriott, F. H. C., & Wu, E. K. H. (1993). Bandwidth selection in robust smoothing. *Journal of Nonparametric Statistics, 2*(4), 333-339. doi:10.1080/10485259308832562
20. Lijian, Y., & Rolf, T. (1999). Multiverait bendwidth selection for local linear regression. *Statist, 61*, 793-815.
21. Müller, H.-G. (1988). Nonparametric regression analysis of longitudinal data.
22. Mustafa, M. Y., & Algamal, Z. Y. (2021). Smoothing parameter selection in kernel nonparametric regression using bat optimization algorithm. *Journal of Physics: Conference Series, 1897*(1). doi:10.1088/1742-6596/1897/1/012010
23. Nychka, D. (1991). Choosing a range for the amount of smoothing in nonparametric regression. *Journal of the American Statistical Association, 86*(415), 653-664.
24. Opsomer, J. D., & Miller, C. P. (2007). Selecting the amount of smoothing in nonparametric regression estimation for complex surveys. *Journal of Nonparametric Statistics, 17*(5), 593-611. doi:10.1080/10485250500054642
25. Qiu, H., & Duan, H. (2020). A multi-objective pigeon-inspired optimization approach to UAV distributed flocking among obstacles. *Information Sciences, 509*, 515-529. doi:10.1016/j.ins.2018.06.061
26. Rice, J. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics, 12*(4), 1215-1230.
27. Schafer, C., & Wasserman, L. (2013). Tutorial on Nonparametric Inference: Carnegie Mellon University.
28. Schucany, W. R. (1995). Adaptive Bandwidth Choice for Kernel Regression. *Journal of the American Statistical Association, 90*(430), 535-540. doi:10.1080/01621459.1995.10476545
29. Shang, H. L., Zhang, X., & Shang, M. H. L. (2014). Package ‘bbemkr’.
30. Sushnigdha, G., & Joshi, A. (2018). Re-entry trajectory optimization using pigeon inspired optimization based control profiles. *Advances in Space Research, 62*(11), 3170-3186. doi:10.1016/j.asr.2018.08.009

31. Yan, L., Qu, B., Zhu, Y., Qiao, B., & Suganthan, P. N. (2019). Dynamic economic emission dispatch based on multi-objective pigeon-inspired optimization with double disturbance. *Science China Information Sciences*, *62*(7). doi:10.1007/s11432-018-9715-2
32. Yang, Z., Duan, H., Fan, Y., & Deng, Y. (2018). Automatic Carrier Landing System multilayer parameter design based on Cauchy Mutation Pigeon-Inspired Optimization. *Aerospace Science and Technology, 79*, 518-530. doi:10.1016/j.ast.2018.06.013
33. Zhang, Z. G., Chan, S. C., Ho, K. L., & Ho, K. C. (2008). On Bandwidth Selection in Local Polynomial Regression Analysis and Its Application to Multi-resolution Analysis of Non-uniform Data. *Journal of Signal Processing Systems, 52*(3), 263-280. doi:10.1007/s11265-007-0156-4
34. Zhong, Y., Wang, L., Lin, M., & Zhang, H. (2019). Discrete pigeon-inspired optimization algorithm with Metropolis acceptance criterion for large-scale traveling salesman problem. *Swarm and Evolutionary Computation, 48*, 134-144. doi:10.1016/j.swevo.2019.04.002
35. Zhou, H., & Huang, X. (2018). Bandwidth selection for nonparametric modal regression. *Communications in Statistics - Simulation and Computation, 48*(4), 968-984. doi:10.1080/03610918.2017.1402044
36. Żychaluk, K. (2014). Bootstrap bandwidth selection method for local linear estimator in exponential family models. *Journal of Nonparametric Statistics, 26*(2), 305-319. doi:10.1080/10485252.2014.885023