

**Smoothing parameter selection in Nadaraya-Watson kernel nonparametric regression using nature-inspired algorithm optimization**

Zinah Ameer Basheer \*  
[azozdodoserzen@gmail.com](mailto:azozdodoserzen@gmail.com)

Dr Zakariya Yahya AlgamaL\*\*  
[zakariya.algamaL@uomosul.edu.iq](mailto:zakariya.algamaL@uomosul.edu.iq)

---

**Abstract**

In the context of Nadaraya-Watson kernel nonparametric regression, the curve estimation is fully depending on the smoothing parameter. At this point, the nature-inspired algorithms can be used as an alternative tool to find the optimal selection. In this paper, a firefly optimization algorithm method is proposed to choose the smoothing parameter in Nadaraya-Watson kernel nonparametric regression. The proposed method will efficiently help to find the best smoothing parameter with a high prediction. The proposed method is compared with four famous methods. The experimental results comprehensively demonstrate the superiority of the proposed method in terms of prediction capability.

**Keywords:** Nadaraya-Watson estimator; firefly optimization algorithm; smoothing parameter selection.

This is an open access article under the CC BY 4.0 license  
<http://creativecommons.org/licenses/by/4.0/>

**1. Introduction**

The nonparametric regression model (NRM) and methods of estimation have been developed mainly in the last years (Ali, 2019; Hazelton & Cox, 2016). Kernel regression estimates are one of the most popular nonparametric estimates. In a univariate case, these estimates depend on a bandwidth, which is a smoothing parameter controlling smoothness of an estimated curve and a kernel which is considered as a weight function (Schucany, 1995; Slaoui, 2016; Steland, 2012).

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; Nychka, 1991; Opsomer & Miller, 2007; Rice, 1984; Schucany, 1995; Zhang, Chan, Ho, & Ho, 2008; Zhou & Huang, 2018; Żychaluk, 2014).

---

\* Researcher / College of Computer science and Mathematics / Mosul University.

\*\*Department of Statistics and Informatics, University of Mosul, Mosul, Iraq.

In this paper, a firefly optimization algorithm method, which is a natural-inspired continuous algorithm, is proposed to choose smoothing parameter in Nadaraya-Watson kernel nonparametric regression. The proposed method will efficiently help to find the best smoothing parameter with a high prediction. The superiority of the proposed method in different simulated examples and a real data application is proved.

This paper is organized as follows. The description of the Nadaraya-Watson kernel nonparametric regression and the smoothing parameter selection are covered in Section 2. The details of the firefly optimization algorithm are covered in Section 3. Section 4 contains the details of our proposed method. The illustration of the proposed method through simulation studies and through real data application is given in Sections 5 and 6. In section 7, the conclusion is covered.

## 2. Smoothing parameter selection

The nonparametric regression model, often estimated by estimators of the Nadaraya-Watson type, forms an attractive framework for diverse areas such as engineering, econometrics, environmetrics, social sciences, and biometrics (Steland, 2012).

In NRM, we have a set of univariate observations  $\{(x_i, y_i)\}_{i=1}^n \in \mathbb{R}$  and the model can be defined as:

$$y_i = m(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

where  $m(x_i)$  is the function of unknown regression and  $\varepsilon_i$  are random errors with mean equal to zero and variance  $\sigma_\varepsilon^2$ . The nonparametric regression depend on weighted mean of the dependent variable, the weights are the distance between the observations of independent variable measured by a smoothing parameter.

One of the techniques nonparametric regression estimate is the Nadaraya-Watson (NW) kernel function estimator which is more flexible than the other nonparametric techniques and it provides an accurate predictor of observations (Ali, 2019; Kyung Lee, Park, & Su Park, 2007; Li & Palta, 2009).

The kernel estimator of  $f(x)$  at the point  $x$ , in general, is defined as

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \quad (2)$$

where  $K$  represents kernel probability density function centered at each point  $x_i$ , and the smoothing parameter  $h > 0$  is known as fixed bandwidth. The NW kernel function estimator with a fixed  $h$  is defined as the following:

$$\hat{m}_{NW}(x_i) = \frac{\sum_{i=1}^n y_i K\left(\frac{x - x_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)}. \quad (3)$$

The NW kernel estimator depends on smoothing parameter  $h$  (bandwidth). It controls the amount of curve smoothing where large  $h$  value leads to a smooth density estimate (Rice, 1984; Schucany, 1995; Utami, Haris, Prahutama, & Purnomo, 2020; Yoon Kim & Park, 2002).

The optimal bandwidth of the NW kernel estimator is the value which minimized the mean integrated squared error and it is obtained by integration the mean squares of errors (IMSE). There are many different methods to select the value of  $h$ . Among them, in 1981, Friedman and Stuetzle used this strong reparability to identify the components of the nonparametric regression model when  $h$  is unknown, and proposed a kernel-based consistent and asymptotically normal estimator (Friedman & Stuetzle, 1981). In 1982, the researcher Abramson suggested the law of inverse square root to estimate  $h$  in variable kernel density function, which reduced the bias more than fixed  $h$  estimator (Abramson, 1982). In 1986, the researcher Silverman suggested an adaptation for the kernel function estimator by varying the  $h$  as nonparametric estimation depends on geometric mean (Silverman, 1986). In 1987, the researchers Scott and Terrell discussed relationship between the biased and unbiased cross-validation and the variable  $h$  used instead than the fixed  $h$  in the case of long-tail distribution (Scott & Terrell, 1987). There are several authors handling the problem of selection the smoothing parameter, 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 et al., 1993; Nychka, 1991; Opsomer & Miller, 2007; Rice, 1984; Schucany, 1995; Zhang et al., 2008; Zhou & Huang, 2018; Żychaluk, 2014).

### 3. Firefly optimization algorithm

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. Swarm intelligence algorithms have been widely studied and successfully applied to a variety of complex optimization problems. The firefly algorithm (FFA), is one of the recent novel swarm intelligence methods and the most powerful optimization algorithms, which was developed by Yang (2013).

Firefly algorithm has been proved to be a good performance and the effectiveness for solving various optimization problems (Fister, Fister, Yang, & Brest, 2013). The firefly algorithm has been inspired by the simulation of the social behavior of fireflies on the basis of the flashing lights or the flash attractiveness. By representing the advantage of some flashing characteristics of fireflies and how fireflies interact with flashing lights, the firefly flash is a signal system which used to attract another firefly (Yelghi & Köse, 2018).

For simplicity in the description and modeling of the FFA, Yang (2013) formulated the following three idealized important rules for the protocol to prefer the search process in FFA. The protocol standardizes certain firefly characteristics and describes the behavior of artificial fireflies as follows:

- (1) All fireflies are unisex, implying that all the fireflies of a population can attract each other regardless of their sex.
- (2) The degree of the attractiveness of a firefly is proportional to its brightness, therefore for any two flashing fireflies, the less bright one will move towards the brighter firefly and the more brightness means the less distance between two fireflies. Attractiveness is proportional to the brightness which decreases as the distance increase between tow fireflies. If there is no brighter firefly than a particular firefly, it will move randomly.
- (3) The brightness of each firefly is affected or determined by the value of the objective function to be optimized. In the maximization problem, the brightness of each firefly can be directly proportional to the value of the fitness function. In the case of a minimization problem, the brightness of each firefly is inversely proportional to the value of objective function.

In the implementation of the FFA, each member is called firefly in the swarm. Each firefly represents a candidate solution in the dimensional search space. The brighter locations are assumed to represent better solutions. Then, the algorithm tries to help fireflies to find these locations in the search space. The firefly's attractiveness is determined by its brightness,

which in turn is associated with the objective function for a given optimization problem. The brightness decreases when the distance between a firefly and the target location increases. The attraction between fireflies is based on the differences in brightness. This means that a less bright can move to a brighter firefly by the attraction. If none of the fireflies are brighter than a particular firefly, it will move randomly. During the search process and because of the attractions among fireflies, fireflies can move towards new locations or positions through the attraction and find new candidate solutions.

Mathematically, assume that there are  $n_f$  of fireflies in the swarm (populations size) are randomly distributed in the  $D$ -dimensional search space. During the evolutionary process, each firefly has a position vector denoted as  $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{id}\}$ , where  $i = 1, 2, \dots, n_f$  and  $d \in D$  is the dimensionality of the solutions.

The distance between any two fireflies  $i$  and  $j$ , at positions  $x_i$  and  $x_j$  in the search space, respectively, is the Cartesian distance which can be calculated using the following equation

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{d=1}^D (x_{id} - x_{jd})^2}. \quad (4)$$

Each firefly has its light intensity or brightness. The brightness value is used to evaluate the goodness of firefly, which is affected by the landscape of the optimization problem. The brightness of firefly  $i$  at a particular or current position  $x$  can be denoted by the objective function value as follows:

$$I(x_i) = f(x_i) \quad (5)$$

The light intensity of the firefly is directly proportional to its brightness and is related to objective values. In comparing the two fireflies, both fireflies are attracted, the firefly which has a lower light intensity is attracted toward the other firefly with the higher light intensity. The light intensity of a firefly depends on the intensity  $I_0$  of light emitted by a firefly and the distance  $r_{ij}$  between two fireflies. Light intensity  $I(r)$  can be described by a monotonically decreasing function of the  $r_{ij}$  which can be formulated as follows:

$$I(r) = I_0 e^{-\gamma r^2}, \quad (6)$$

where  $\gamma$  is used to control the decrease of the light intensity or brightness and can be taken as a constant.

Each firefly has its distinctive attractiveness which indicates how powerful it attracts other members in the swarm. Attractiveness,  $\beta$ , is relative, which means that it must be judged by others, and therefore varies with the distance  $r_{ij}$ . As mentioned earlier, the brightness decreases with the distance from the source and the light is also absorbed by the air, therefore the attractiveness must be allowed to vary with differing degrees of absorption (Karthikeyan, Asokan, & Nickolas, 2014). Thus, the main form of the attractiveness of a firefly is defined as the following equation:

$$\beta(r) = \beta_0 e^{-\gamma r^2}, \quad (7)$$

where  $\beta(r)$  represents attractiveness function of a firefly at a distance,  $r$ , and  $\beta_0$  denotes the initial attractiveness of a firefly at distance  $r=0$  and it can be constant. For implementation usually  $\beta_0$  set to be 1 for most problems.

The fireflies will try to move to the best position. This means that the lower light intensity one will be attracted by the brighter one. The location updates for each pair of fireflies  $i$  and  $j$ . For each firefly  $x_i$ , it is compared to other all fireflies  $x_j, j=1,2,\dots,n_f$ . If firefly  $j$  at position  $x_j$  is brighter than firefly  $i$ , then  $x_i$  will move towards  $x_j$  by the attraction. The movement is defined as:

$$x_{id}^{(t+1)} = x_{id}^{(t)} + \beta_0 e^{-\gamma r^2} (x_{jd}^{(t)} - x_{id}^{(t)}) + \alpha_i \varepsilon_{id}^{(t)}, \quad (8)$$

where  $\alpha_i$  is the randomization parameter,  $\gamma$  is an absorption coefficient which controls the decrease of the light intensity, and  $\varepsilon_{id}^{(t)} = (rand - 0.5)$ , where  $rand$  is a random number from uniform distribution with  $[0, 1]$ .

#### 4. The proposed method

The efficiency of NW kernel estimator largely depends on an appropriately choosing the smoothing parameter,  $h$ . As a result, it is of crucial importance selecting a suitable value of the  $h$ . In literature, the most widely used method for selecting  $h$  is the cross-validation (CV), which is a data-driven approach (Chen, 2015; Scott & Terrell, 1987).

In this paper, a FFA is proposed to determine the smoothing parameter in the NW 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 fireflies, is set to 25 and the number of iterations is  $t_{\max}=100$ .
- (2) The positions of each firefly are randomly determined. The position of a firefly represents the smoothing parameter,  $h$ . The initial positions of the fireflies are generated from a uniform distribution within the range  $[0,10]$ .
- (3) The fitness function is defined as
 
$$\text{fitness} = \min \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (9)$$
- (4) The positions are updated using Eq. (8).
- (5) Steps 3 and 4 are repeated until a  $t_{\max}$  is reached.

## 5. 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, AIC, and plug-in method (PM) are also conducted. Three sizes of samples are taken as:  $n = 50, 100, 150$ . In addition, the type of kernel is setting as Epanechnikov kernel type.

**Case 1:** In the case, we use the regression function

$$y_i = x^3(1-x)^3 + \varepsilon_i \sim N(0, 0.003) \quad (10)$$

The explanatory variable,  $x$  is generated from uniform distribution with the range 0 and 1.

**Case 2:** In the case, we use the regression function

$$y_i = \sin(\pi x) \cos(3\pi x^5) + \varepsilon_i \sim N(0, 0.05) \quad (11)$$

The explanatory variable,  $x$  is generated from uniform distribution with the range 0 and 1.

**Case 3:** In the case, we use the regression function

$$y_i = x \sin(2\pi x) + \varepsilon_i \sim N(0, 0.05) \quad (12)$$

The explanatory variable,  $x$  is generated from uniform distribution with the range 0 and 1.

**Case 4:** In the case, we use the regression function

$$y_i = \sin(2\pi x) + \varepsilon_i \sim N(0, 0.5) \quad (13)$$

The explanatory variable,  $x$  is generated from uniform distribution with the range 0 and 1.

**Case 5:** In the case, we use the regression function

$$y_i = \sin(4x) + \varepsilon_i \sim N(0, 0.11) \quad (14)$$

The explanatory variable,  $x$  is 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.

The results presented in Tables 1 – 5 show that the FFA algorithm yielded higher prediction accuracy than did CV, GCV, AIC, and PM methods for the case 1, case 2, case 3, case 4, and case 5. The proposed method, FFA, has superior reduction in terms of IMSE. For example, for case 2, the reduction in IMSE using FFA was 5.52%, 5.02%, 5.97%, and 7.84% compared with CV, GCV, AIC, and PM methods, respectively.

With respect to the sample size, there is a decreasing in the IMSE values when the sample size increasing. However, the performance of the proposed method, FFA, is still the best among others in all cases.

Table 1: The average IMSE for the case 1

Methods	n=50	n=100	n=150
CV	0.2361	0.2175	0.2168
GCV	0.2365	0.2532	0.2321
AIC	0.3974	0.3685	0.3366
PM	0.6224	0.4510	0.4142
FFA	<b>0.1441</b>	<b>0.1422</b>	<b>0.1107</b>

Table 2: The average IMSE for the case 2

Methods	n=50	n=100	n=150
CV	0.2673	0.1312	0.1216
GCV	0.2651	0.1468	0.1316
AIC	0.3203	0.2994	0.2596
PM	0.5758	0.4786	0.4152
FFA	<b>0.1311</b>	<b>0.1108</b>	<b>0.1138</b>

Table 3: The average IMSE for the case 3

Methods	n=50	n=100	n=150
CV	0.2772	0.2502	0.2485
GCV	0.2606	0.2402	0.2385
AIC	0.2914	0.2865	0.2493
PM	0.4254	0.39567	0.3891
FFA	<b>0.1781</b>	<b>0.1628</b>	<b>0.1381</b>

Table 4: The average IMSE for the case 4

Methods	n=50	n=100	n=150
CV	0.3681	0.2916	0.3524
GCV	0.3587	0.2886	0.3343
AIC	0.3862	0.3572	0.3061
PM	0.4154	0.4106	0.4251
FFA	<b>0.2528</b>	<b>0.2311</b>	<b>0.2101</b>

Table 5: The average IMSE for the case 5

Methods	n=50	n=100	n=150
CV	0.0671	0.0627	0.0572
GCV	0.0558	0.0468	0.0406
AIC	0.0624	0.0512	0.0424
PM	0.1256	0.1124	0.1056
FFA	<b>0.0315</b>	<b>0.0308</b>	<b>0.0227</b>

## 6. Real application results

To investigate the performance of our proposed method in real application, 255 observations are daily taken from 2ed January/2019 to 13ed March/2020). This data is related to ASIACell Communication Company in

Iraq. The response variable is the number of deals and the explanatory variable is the closed price of the share in Iraqi dinar. The estimated IMSE values and the value of the smoothing parameter,  $h$ , is reported in Table 6. It is clearly seen from Table 6, that the proposed method, FFA, achieved the lowest prediction error. It reduced the IMSE by 5.08%, 4.87%, 3.64%, and 3.67% than those obtained by CV, GCV, AIC, and PM, respectively. Figure 1 shows the smoothing curve of the used methods in addition to the linear model which is estimated by ordinary least square method (OLS). It is clearly seen that our proposed method give a more smooth curve than the others.

Table 6: The IMSE results of the application data

Methods	$h$	IMSE
CV	1.970	0.8547
GCV	1.851	0.7741
AIC	1.024	0.6526
PM	1.051	0.6628
FFA	<b>0.508</b>	<b>0.4088</b>

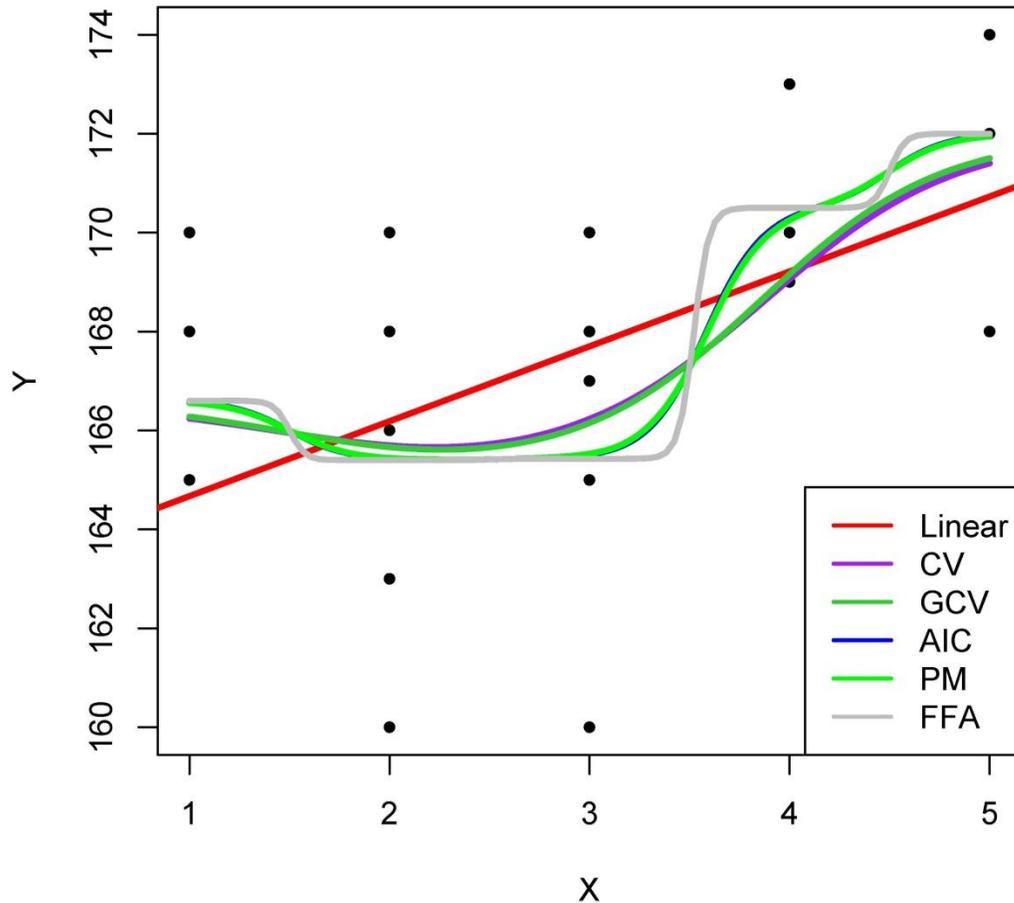


Figure 1: The estimated smoothing curve for the real data application over the used methods with the OLS method.

## 7. Conclusion

In this paper, the problem of selecting smoothing parameter in Nadaraya-Watson kernel nonparametric regression is considered. A firefly optimization algorithm was proposed to choose the parameter of smoothing parameter. The results obtained from simulation and real data application demonstrated the superiority of the proposed method, FFA, in terms of IMSE comparing with other competitor methods.

**REFERENCES**

- Abramson, I. S. (1982). On bandwidth variation in kernel estimates-a square root law. *The Annals of Statistics*, 1217-1223.
- 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
- Chen, S. (2015). Optimal Bandwidth Selection for Kernel Density Functionals Estimation. *Journal of Probability and Statistics*, 2015, 1-21. doi:10.1155/2015/242683
- Chu, C.-K., & Marron, J. S. (1991). Comparison of two bandwidth selectors with dependent errors. *The Annals of Statistics*, 19(4), 1906-1918.
- Chu, C. (1995). Bandwidth selection in nonparametric regression with general errors. *Journal of statistical planning and inference*, 44(3), 265-275.
- 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
- 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
- Fister, I., Fister, I., Yang, X.-S., & Brest, J. (2013). A comprehensive review of firefly algorithms. *Swarm and Evolutionary Computation*, 13, 34-46. doi:10.1016/j.swevo.2013.06.001
- 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.
- Friedman, J. H., & Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association*, 76(376), 817-823.
- 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
- Hazelton, M. L., & Cox, M. P. (2016). Bandwidth selection for kernel log-density estimation. *Computational Statistics & Data Analysis*, 103, 56-67. doi:10.1016/j.csda.2016.05.003
- Karthikeyan, S., Asokan, P., & Nickolas, S. (2014). A hybrid discrete firefly algorithm for multi-objective flexible job shop scheduling problem with limited resource constraints. *The International Journal of Advanced Manufacturing Technology*, 72(9-12), 1567-1579. doi:10.1007/s00170-014-5753-3

- 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
- Koláček, J., & Horová, I. (2017). Bandwidth matrix selectors for kernel regression. *Computational Statistics*, 32(3), 1027-1046. doi:10.1007/s00180-017-0709-3
- Kyung Lee, Y., Park, B. U., & Su Park, M. (2007). A simple and effective bandwidth selector for local polynomial quasi-likelihood regression. *Journal of Nonparametric Statistics*, 19(6-8), 255-267. doi:10.1080/10485250701761086
- Lee, T. C., & Solo, V. (1999). Bandwidth selection for local linear regression: a simulation study. *Computational Statistics*, 14(4), 515-532.
- 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
- Li, J., & Palta, M. (2009). Bandwidth selection through cross-validation for semi-parametric varying-coefficient partially linear models. *Journal of Statistical Computation and Simulation*, 79(11), 1277-1286. doi:10.1080/00949650802260071
- Nychka, D. (1991). Choosing a range for the amount of smoothing in nonparametric regression. *Journal of the American Statistical Association*, 86(415), 653-664.
- 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
- Rice, J. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics*, 12(4), 1215-1230.
- 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
- Scott, D. W., & Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82(400), 1131-1146.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26): CRC press.
- Slaoui, Y. (2016). Optimal bandwidth selection for semi-recursive kernel regression estimators. *Statistics and Its Interface*, 9(3), 375-388. doi:10.4310/SII.2016.v9.n3.a11

- Steland, A. (2012). Sequential Data-Adaptive Bandwidth Selection by Cross-Validation for Nonparametric Prediction. *Communications in Statistics - Simulation and Computation*, 41(7), 1195-1219. doi:10.1080/03610918.2012.625853
- Utami, T. W., Haris, M. A., Prahutama, A., & Purnomo, E. A. (2020). Optimal knot selection in spline regression using unbiased risk and generalized cross validation methods. *Journal of Physics: Conference Series*, 1446, 012049. doi:10.1088/1742-6596/1446/1/012049
- Yang, X.-S. (2013). Multiobjective firefly algorithm for continuous optimization. *Engineering with Computers*, 29(2), 175-184.
- Yelghi, A., & Köse, C. (2018). A modified firefly algorithm for global minimum optimization. *Applied Soft Computing*, 62, 29-44. doi:10.1016/j.asoc.2017.10.032
- Yoon Kim, T., & Park, C. (2002). Regression Smoothing Parameter Selection Using Cross Residuals Sum. *Communications in Statistics - Theory and Methods*, 31(12), 2275-2287. doi:10.1081/sta-120017225
- 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
- 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
- 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