researchvision.us

Symmetria: Journal of Advanced Physics and Mathematical Review
Vol. 2 | No. 1 | 2025 pp. 1-8 | **ISSN: 3066-3636**

Comparison Some Nonparametric Estimation Methods in Spatial Statistics

Jaufar Mousa Mohammed[1]

[1]Department of Statistics Sciences, College of Administration & Economics, Kirkuk University, Iraq

Submitted: 30-Aug, 2025
Accepted: 29-Sept, 2025
Published: 12-Oct, 2025

Vol. 2, **No.** 1, 2025.
10.62762/SJAPMR.2025.000000

***Corresponding author:**
Jaufar Mousa Mohammed
jaufar@uokirkuk.edu.iq

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

In this study, we systematically compare four nonparametric approaches in spatial statistics: Local Polynomial (LP), Radial Basis Function (RBF) and its normalized version, Normalized RBF and Thin Plate Splines (TPS). These methods were tested using data of traffic accidents for the year 2023 in the streets network of 15 Iraqi provinces, to assess their predictive ability in modeling spatial distributions of accidents. Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are among the main statistical measures used in comparing the methods, in which MSE and RMSE were calculated. We found that the strategies were associated with different levels of smoothing, computational efficiencies and flexibility in representing spatial patterns, with Thin Plate Splines striking a good balance between accuracy and flexibility. These results highlight the need for careful method selection that reflects data properties and spatial analysis objectives in order to improve understanding of spatial phenomena and ensure data-driven decisions through sound statistical model selection.

Key words: Local estimation , Splines , Radial Basis Function

Introduction

In spatial statistics, the estimation of spatial surfaces and trends are fundamental to discovering how data (any social, environmental or geographical data) vary in space. As a consequence of nonlinear complexities and uneven spatial variation during the process of modeling spatial phenomenon, researchers generally encounter various problems which are growing concerns, leading to a list of nonparametric techniques that do not have an existing model form (Boyle et al., [1]). Of these methods, the LPR is a special case of nonparametric method for estimating spatial surfaces. It is an approach that estimates the polynomial models locally on neighborhoods, which endows it with great flexibility at capturing small spatial variations [2]. Alternatively, RBF (Radial Basis Functions) have been applied to multivariate surface estimation because they are capable of generating smooth applicable models for radial distances between data points [3]. Normalized versions of these functions (Normalized RBF) were also designed in order to stabilize the estimate and diminish outliers effect [4]. Another popular one of the spatial smoothing techniques is the Thin Plate Splines (TPS) method. This model is based on a minimization of the bending energy of an estimated surface, leading to a smooth representation that approximates spatial distributions and takes into account geometrical measures of the studied regions [5]. The objective of this paper is to compare the four methods in the context of spatial statistics and to

address how they perform in terms of accuracy and capturing spatial patterns, on both simulated and real data. This comparison will hopefully offer a useful reference for deciding a proper estimation method, based on the data type and statistical purposes [7].

Materials and Methods

Research Objective

The purpose of this paper is a numerical scientific comparison among four nonparametric estimation methods used in spatial statistics, and they are: Local Polynomial, Radial Basis Function (RBF) and its normalized form (Normalized RBF), Thin Plate Splines (TPS). The comparison is made in terms of estimation accuracy, smoothing fidelity, computational efficiency and modeling flexibility for spatial correlation patterns.

Local Polynomial Method

The local polynomial estimation is analogous to TSA, but instead of the whole dataset it uses a moving window data. That is, the surface constructed by the set of polynomial functions and its local trends and weights are suitable. We use the notation for the estimator, where one estimate is determined at each location. Weighted least squares is used to decrease the estimating error.

$\sum_{i=1}^{n} \lambda_i [\hat{Z}(x_i, y_i) - Z_0(x_i, y_i)]^2 \dots\dots\dots\dots (1)$

Where $\hat{Z}(x_i, y_i)$is the estimated value of the attribute at location (x, y), and n is the number of data points within the local window. Here, $\lambda_i$is computed using the following equation:

$\lambda_i = \exp(-3d_{i0}/h) \dots\dots\dots\dots (2)$

Where $d_{i0}$is the distance between the data points and the center of the window, and h is a parameter that controls the decay of weights with distance. Finally, $Z_0(x, y)$ represents the value of the polynomial.

A first-order polynomial takes the following form:

$Z_0(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i \dots\dots\dots (3)$

And a second-order polynomial takes the following form:

$Z_0(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 x_i^2 + \beta_4 y_i^2 + \beta_5 x_i y_i \dots\dots\dots (4)$

The coefficients $\beta_i$are minimized in order to re-estimate them at the central point. As the window moves, the local polynomial estimation is updated accordingly. This results in a piecewise and moderately smooth estimation surface. Although not highly precise, it is more flexible than trend surface analysis. However, it involves several parameter choices when incorporating existing data. In cases where short-range variations are present, local polynomial estimation can be a useful method for capturing such local structures.

Radial Basis Function

Let $\hat{Z}_0 = \hat{Z}(r_0)$ be the linear estimator at location r_0, based on N observations $Z(r_1), Z(r_2), \dots\dots Z(r_N)$ at locations $r_1, r_2, \dots\dots\dots r_N$.It is assumed that each of these observations has a similar influence within the surrounding area and in all directions. These effects can be represented by a function ϕ, which is a radial basis function (RBF) depending on the distance $h = \|r_0 - r_i\|$ between the target location r_0 and each observed location r_i.Thus, the function is expressed as: $\phi = \phi(\|r_0 - r_i\|$, Accordingly, the estimated value atr_0 is given by the following form:

$\hat{Z}_0 = \sum_{i=1}^{N} \lambda_i \phi(\|r_0 - r_i\|) \dots\dots\dots\dots (5)$

Where λ_i is the weight associated with location i .Assume that the weights λ_i are chosen such that the estimator yields an exact value at the known point r_j.

$\hat{Z}_j = \sum_{i=1}^{N} \lambda_i \phi(\|r_j - r_i\|) \quad , \quad j = 1,2,3, \ldots \ldots N$ ……(6)

This condition generates a system of N equations with N unknowns, which can be written in matrix form as follows:

$$\begin{bmatrix} Z_1 \\ Z_2 \\ . \\ . \\ Z_N \end{bmatrix} = \begin{bmatrix} \phi(\|r_1 - r_1\| & \phi(\|r_1 - r_2\| & . & . & \phi(\|r_1 - r_N\| \\ \phi(\|r_2 - r_1\| & \phi(\|r_2 - r_2\| & . & . & \phi(\|r_2 - r_N\| \\ . & . & . & . & . \\ . & . & . & . & . \\ \phi(\|r_N - r_1\| & \phi(\|r_N - r_2\| & . & . & \phi(\|r_N - r_N\| \end{bmatrix} * \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ . \\ . \\ \lambda_N \end{bmatrix}$$

Or

$Z = \lambda\phi$……..(7)

Often polynomial term

$P(r_j) = c_0 + c_1 x_j + c_2 y_j + \cdots \ldots . = [1 \quad x_j \quad y_j \quad . \quad . \quad . \quad .] * c$

Where

$c = [c_0 \quad c_1 \quad c_2 \quad . \quad . \quad . \quad .]^t$

The polynomial coefficient vector is added to the estimator at location $r_j = [x_j \quad y_j]^t$ which is two-dimensional,

$\hat{Z}_j = P(r_j) + \sum_{i=1}^{N} \lambda_i \phi(\|r_j - r_i\|)$ ……(8)

Where $P^t \lambda = 0$ and

$$P = \begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & y_1^2 & x_1 y_1 & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ 1 & x_N & y_N & x_N^2 & y_N^2 & x_N y_N & . & . & . & . \end{bmatrix}$$

The applied polynomial is often defined such that the constant term corresponds to one of its columns, or to the linear polynomial used.

To solve the system of equations $Z = \lambda\phi$ under the constraint $P^t \lambda = 0$, we obtain:

$$\begin{bmatrix} Z \\ 0 \end{bmatrix} = \begin{bmatrix} \phi & P \\ P^t & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ c \end{bmatrix} \ldots \ldots \ldots \ldots \ldots \ldots . (9)$$

(similar to ordinary kriging with a Lagrange multiplier), and the degree of the polynomial may depend on the chosen RBF.

Radial Basis Function Normalize

The NRBF estimator is derived from:

$$\hat{Z}_0 = \frac{\sum_{i=1}^{N} \lambda_i \phi(\|r_0 - r_i\|)}{\sum_{i=1}^{N} \phi(\|r_0 - r_i\|)} \ldots \ldots \ldots \ldots \ldots \ldots \ldots (10)$$

which leads to

$Z_j \sum_{i=1}^{N} \phi(\|r_j - r_i\|) = \sum_{i=1}^{N} \lambda_i \phi(\|r_j - r_i\|) \quad , \quad j = 1,2,3, \ldots \ldots . . N$

For the weights, in matrix form, this leads to the same system of equations as above, but with the elements in the right-hand side vector changing from Z_j to:

$Z_j \sum_{i=1}^{N} \phi(\|r_j - r_i\|)$

Thin Plate Splines

Let v_i denote the target function values at locations (x_i, y_i) in the plane, with $i = 1, 2, \ldots, p$. In particular, we will set v_i equal to the target coordinates (x'_i, y'_i) in turn to obtain one continuous transformation for each coordinate. We assume that the locations (x_i, y_i) are all different and are not collinear. The TPS interplant $f(x, y)$ minimizes the bending energy

$$Jf = \iint \left[\left(\frac{\vartheta f^2}{\vartheta x^2} \right)^2 + 2 * \left(\frac{\vartheta f^2}{\vartheta x \vartheta y} \right)^2 + \left(\frac{\vartheta f^2}{\vartheta y^2} \right)^2 \right] dxdy$$

and has the form

$$f(x,y) = a_1 + a_x x + a_y y + \sum_{i=1}^{p} w_i U(\|(x_i, y_i) - (x,y)\|) \dots\dots\dots (11)$$

Where

$r = U(r)$

$r^2 \log(r) = \sqrt{(x - x_i)^2 + (y - y_i)^2} \dots\dots\dots\dots (12)$

In order for $f(x,y)$ to have square integrable second derivatives, we require that

$\sum_{i=1}^{p} w_i = 0$ and $\sum_{i=1}^{p} w_i x_i = \sum_{i=1}^{p} w_i y_i = 0$

Together with the interpolation conditions, $f(x_i, y_i) = v_i$ this yields a linear system for the TPS coefficients

$$\begin{bmatrix} K & P \\ P' & O \end{bmatrix} \begin{bmatrix} w \\ a \end{bmatrix} = \begin{bmatrix} v \\ 0 \end{bmatrix} \dots\dots\dots\dots\dots\dots (13)$$

where $Kij = U(\|(xi, yi) - (x, y)\|)$, the ith row of P is $(1, xi, yi)$, O is a 3×3 matrix of zeros, o is a 3×1 column vector of zeros, w and v are column vectors formed from w_i and v_i, respectively, and a is the column vector with elements a_1, a_x, a_y. We will denote the $(p + 3) \times (p + 3)$ matrix of this system by L; L is nonsingular. If we denote the upper left $p \times p$ block of L^{-1} by L_p^{-1}, then it can be shown that

$$If \propto v^T L_p^{-1} v = w^T K w$$

When there is noise in the specified values v_i, one may wish to relax the exact interpolation requirement by means of regularization. This is accomplished by minimizing

$H[f] = \sum_{i=1}^{p} (v_i - f(x_i, y_i))^2 + \lambda I_f$

The regularization parameter λ, a positive scalar, controls the amount of smoothing; the limiting case of $\lambda = 0$ reduces to exact interpolation. we can solve for the TPS coefficients in the regularized case by replacing the matrix K by $K + \lambda I$, where I is the $p \times p$ identity matrix.

Application

This paper employed real data relating to traffic accidents in 15 provinces of Iraq as listed in Table (1). The information was taken from the Directorate of Transport and Communications Statistics, year 2023. In this paper, is the number of traffic accident at location xxx with spatial coordinates: denoting East–West axis and denotes North-South axis. A log transformation has been performed on the dependent variables to remove heteroscedasiticity.

Table 1. number of road traffic accidents in Iraq for the year 2023

Governorates	$u(x)$	$v(x)$	$Z(x)$
Nineveh	30	62	**475**
Kirkuk	38	54	**386**
Diyala	40	40	**1086**
Anbar	30	38	**462**
Baghdad	39	39	**753**
Babylon	37	28	**1043**
Karbala	35	30	**618**
Wasit	48	30	**1168**

Salahuddin	33	47	**354**
Najaf	36	25	**858**
AL-Qadisiyah	42	27	**728**
Muthanna	43	20	**355**
Dhi Qar	50	20	**532**
Maysan	55	25	**334**
Basra	**60**	**15**	**1667**

The estimation techniques—Local Polynomial Method (LPM), Radial Basis Function (RBF), Normalized RBF (RBFN), and Thin Plate Splines (TPS)—were applied to this dataset as follows.

Results and Discussion

(LPM) Technique

In this method, a first-order polynomial was used. The estimation in LPM is performed independently at each location, meaning that each point has its own local mathematical formulation and set of coefficients. After performing the estimation procedure, the results were obtained as shown in Table (2).

(RBF) Technique

The non-parametric estimation technique based on RBF was employed to estimate spatio-distribution values. The estimate at any unobserved location was a linear combination of basis functions which are a function of the distance between the estimation point and observed data points. The basis function used here was Gaussian and it is defined as follows:

$\emptyset(r) = \exp(-(\varepsilon(r)^2) \quad , \quad \varepsilon = 1$

where ε is the shape parameter that controls the smoothness of the estimated surface [8]. The procedure was implemented using MATLAB, and the estimated values were compared with the observed data. The computational results are presented in Table (2).

(RBFN) Technique

The Normalized Radial Basis Function (NRBF) technique was used to estimate spatial values at new and unmeasured points. This method is based on the same principles as the RBF approach, but it performs normalization on the estimated values to make the prediction more stable in cases of varying point density [9]. The estimated value at each point x0x_0x0 is computed using equation (10), and the Gaussian function, as previously defined in the RBF technique, was also used. This technique demonstrated good results in terms of stability and smoothness, especially in the presence of clustered data points. The RBFN algorithm was implemented using MATLAB, where a custom function was developed to calculate distances between points, apply the radial basis function, and then perform normalization [10]. The results were included in Table (2)

(TPS) Technique

The Thin Plate Spline (TPS) technique was applied to estimate unmeasured spatial values based on a set of measured data [11]. This technique relies on minimizing the bending of the estimated surface while maintaining consistency with the observed values and is considered a natural extension of quadratic interpolation to two dimensions [12]. The estimation function is represented according to Equation (11), and the estimation results are shown in Table (2).

Table 2. Results of the Nonparametric Estimation

LPM, RBF, RBFN, and TPS

Governorates	$Z(x)$	$\widehat{Z}_{LPM}(x)$	$\widehat{Z}_{RBF}(x)$	$\widehat{Z}_{RBFN}(x)$	$\widehat{Z}_{TPS}(x)$
Nineveh	475	475	475	475	**475**
Kirkuk	386	386.26	386	386	**386**
Diyala	1086	1000	1086	1046.305	**1086**
Anbar	462	457.88	462	462	**462**
Baghdad	753	875.43	753	792.6946	**753**
Babylon	1043	835.41	1043	1042.849	**1043**
Karbala	618	715.76	618	618.1425	**618**
Wasit	1168	1143.7	1168	1168	**1168**
Salahuddin	354	353.09	354	354	**354**
Najaf	858	858.76	858	858.0084	**858**
AL-Qadisiyah	728	806.28	728	728	**728**
Muthanna	355	370.76	355	355	**355**
Dhi Qar	532	517.2	532	532	**532**
Maysan	334	353.31	334	334	**334**
Basra	**1667**	**1667**	**1667**	**1667**	**1667**

Discussion

Because the values estimated by non-parametric estimation methods, such as RBF and TPS, strongly coincided with actual measured data values at training points, statistical evaluation measures included Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2) exhibited different behavior among these methods [13]. This disagreement is due to the fact these methods are exact interpolation models, they reproduce the original values only at the training points. So, if you apply this model to the same points it was trained on, the error will be 0 or very close to 0 and you may get a false sense of confidence about the accuracy of your model. In order to overcome such bias, cross-validation was performed (in particular, the K-Fold technique), splitting data points into several subsets [14]. In turn, one subset is left out for testing after training the model on other subsets. This approach provides the means to evaluate how well the model performs at unmeasured locations and thus quantify its actual predicting and generalising abilities. In the light of this realistic assessment, basic discrepancies in accuracy of estimation were found for the different methods, despite a seemingly identical performance at training points [15]. This emphasizes the need to use accurate and up-to-date evaluation metrics when assessing models, instead of directly comparing performance on the training set (Table 3).

Table 3. Results of the comparison criterion between nonparametric estimates

Method	MSE	RMSE	MAE	R^2
LPM	0.0072	0.0850	0.0538	**0.9693**
RBF	0	0	0	**1**
RBFN	0.0003	0.0159	0.0058	**0.9989**
TPS	0	0	0	**1**

Conclusion

Referring to the results presented regarding the estimation of traffic accidents in 15 Iraqi governorates using the four nonparametric estimation techniques, we conclude the following:

1- The nonparametric models (LPM, RBF, RBFN, TPS) demonstrated a high capability in representing the spatial patterns of traffic accidents, as most methods achieved high coefficients of determination (R^2), indicating an accurate match between the original and estimated values.

2- The RBF and TPS techniques outperformed others based on statistical performance indicators (MSE, RMSE, MAE, R^2), where both achieved excellent estimation with zero error and $R^2 = 1$, indicating a perfect match between the original and estimated data. However, this absolute accuracy may suggest the possibility of overfitting, especially when cross-validation is not employed.

3- The RBFN method showed reasonable performance, achieving results close to ideal with an $R^2 \approx 0.9989$ and very low values of MSE and RMSE. This indicates its ability to estimate with good accuracy and generalization capacity, making it a practical choice for application on new or incomplete data.

4- The Local Polynomial Method (LPM) showed relatively weak performance; although it provided acceptable results ($R^2 = 0.9693$), its relative error was higher compared to other methods. This reflects the limited ability of LPM to capture irregular spatial variability in the data, especially in locations experiencing sharp changes in traffic accident counts.

5- Spatial differences in estimation quality appeared across locations, particularly in Baghdad, Karbala, and Babil, which may indicate spatial heterogeneity requiring more flexible models to capture the spatial structure of the data.

Recommendations

1. Use the most accurate spatial estimation techniques such as RBF or TPS when precise sample estimation models are required, while preferring RBFN in cases needing broader application to out-of-sample data due to its high generalization capability.
2. Integrate cross-validation in evaluation: To ensure result reliability and avoid over fitting, it is recommended to use cross-validation techniques (e.g., K-Fold CV) for objective and comprehensive model performance assessment.
3. Expand into spatio-temporal modeling: Given the dynamic nature of traffic accidents, it is advisable to develop spatio-temporal models that consider temporal changes alongside spatial distribution.
4. Employ nonparametric models in decision-making: Best-performing models (such as RBFN and RBF) can be utilized in traffic authorities' decision support systems to identify high-risk governorates and propose targeted interventions.

5. Expand explanatory variables: It is suggested to incorporate additional variables such as population density, vehicle count, infrastructure condition, and number of intersections within governorates to improve model performance and enhance their predictive capability.

References

[1] M. Avery, *Literature review for local polynomial regression*, unpublished manuscript, 2013.
[2] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, 2002.
[3] M. D. Buhmann, "Radial basis functions," *Acta Numerica*, vol. 9, pp. 1–38, 2000.
[4] N. Cressie, *Statistics for Spatial Data*, Hoboken, NJ, USA: John Wiley & Sons, 2015.
[5] G. Donato and S. Belongie, "Approximate thin plate spline mappings," in *Proc. Eur. Conf. Comput. Vis.*, Berlin, Heidelberg: Springer, 2002.
[6] J. Fan, *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*, London, U.K.: Routledge, 2018.
[7] A. E. Gelfand *et al.*, *Handbook of Spatial Statistics*, Boca Raton, FL, USA: CRC Press, 2010.
[8] M. M. Jaufar, "Determine the best spatial model appropriate to estimate number of deaths from chronic diseases," *J. Math. Probl. Equ. Statist.*, vol. 6, no. 1, pp. 259–265, 2025.
[9] W. Luo, M. Taylor, and S. Parker, "Spatial interpolation for wind data in England and Wales," Intermet, U.K., 2013.
[10] J. M. Mohammed, "Spatial regression analysis using Poisson regression: Applications in studying traffic accidents," *Eur. J. Appl. Sci. Eng. Technol.*, vol. 3, no. 3, pp. 268–275, 2025.
[11] J. M. Mohammed, "Comparing the effectiveness of linear spatial regression and neural networks for spatial data estimation," *Scientia. Technol. Sci. Soc.*, vol. 2, no. 6, pp. 15–25, 2025.
[12] M. J. D. Powell, "The theory of radial basis function approximation in 1990," *Adv. Numer. Anal.*, pp. 105–210, 1992.
[13] G. Sharma and M. Q. Shaw, "Thin-plate splines for printer data interpolation," in *Proc. 14th Eur. Signal Process. Conf.*, 2006.
[14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.
[15] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Cambridge, MA, USA: MIT Press, 2006.