Decision Science Letters 12 (2023) 721–728

Contents lists available at GrowingScience

Decision Science Letters

homepage: www.GrowingScience.com/dsl

A hybrid model of spatial autoregressive-multivariate adaptive generalized Poisson regression spline

Septia Devi Prihastuti Yasmirullah[a,b], Bambang Widjanarko Otok[a*], Jerry Dwi Trijoyo Purnomo[a] and Dedy Dwi Prastyo[a]

[a]*Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[b]*Data Science Technology Study Program, Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga, Surabaya, Indonesia*

C H R O N I C L E

Article history:
Received: May 12, 2023
Received in revised format: June 12, 2023
Accepted: July 16, 2023
Available online: July 16, 2023

Keywords:
Count data
Generalized Poisson
Health policies
MARS
SAR

A B S T R A C T

Several Multivariate Adaptive Regression Spline (MARS) approaches are available to model categorical and numerical (especially continuous) data. Currently, there are other numerical data types—discrete or count data—that call for specific consideration in modeling. Additionally, spatially correlated count data is frequently observed. This has been seen in the case of health data, for example, the number of newborn fatalities, tuberculosis patients, hospital visitors, etc. However, currently no structurally consistent nonparametric regression and MARS model for count data incorporating spatial lag autocorrelation. The SAR-MAGPRS estimator (Spatial Autoregressive - Multivariate Adaptive Generalized Poisson Regression Spline) is developed to fill this gap. Although it can be applied to different count distributions, the estimator was developed in this study under the assumption of a Generalized Poisson distribution. This paper provides an information-theoretic framework for incorporating knowledge of the spatial structure and non-parametric regression models, especially MARS for the count data types. Moreover, the proposed method can assist in modeling the number of diseases while health policies are being developed. The framework presents an application of the Penalized Least Square (PLS) method to estimate the SAR – MAGPRS model.

1. Introduction

A non-parametric regression tool called MARS uses an adaptive regression spline technique to solve the multiple regression problem (Friedman, 1991; Kooperberg, 2014). MARS employs an adaptive algorithm called stepwise, which consists of two processes, i.e., forward and backward. Forward stepwise will increase the basis function until it reaches the maximum number, while backward stepwise will eliminate the basis function, which does not significantly affect the response variable in the model based on the minimum GCV value criterion (Friedman, 1991). MARS has an advantage over other nonparametric regression techniques in that it can efficiently handle many predictor variables, many of which have a nonlinear relationship to the response variable. While the model selection method employs regression splines as a basis function within the least-squares framework (Stoklosa & Warton, 2018), it can also handle multidirectional interactions with flexibility. According to supplementary research, the MARS approach proved effective in modeling the nonlinear relationship between many variables with multicollinearity and high-level interaction (Raj & Gharineiat, 2021). The original MARS commonly used for prediction in data with continuous (numerical) or categorical responses (Friedman, 1991). It is presently widely applied in a wide range of applications, including computer science, medical research, geoscience, engineering, science, etc. (Ampulembang et al., 2015; Dey & Das, 2016; Liu et al., 2019; Otok et al., 2020; Wang et al., 2021; Yasmirullah, Otok, Purnlmo, et al., 2021; Yasmirullah, Otok, Purnomo, et al., 2021; York et al., 2006; Zheng et al.,

* Corresponding author.
E-mail address: bambang_wo@statistika.its.ac.id (B. W. Otok)

doi: 10.5267/dsl.2023.7.004

2019). Meanwhile, when conducting empirical research, researchers frequently extract information that results in small positive integer values — count outcomes. Examples include the number of newborn fatalities, tuberculosis patients, hospital visitors. Then, it categorized the count data types.

We are motivated by the type of discrete or count data with spatial correlation. In an empirical study, the count data type is usually over-dispersed or under-dispersed and possibly spatially correlated across responses. For example, when observing the number of tuberculosis cases in a regency, there may be a spatial correlation between data from the sub-regency (Makalew, Kuntoro, et al., 2019; Makalew, Otok, et al., 2019). One major drawback of the original MARS is that it was not suited for such a situation, the original MARS was created for independent Gaussian responses. Additionally, compared to a standard cross-sectional study, the theory and implementation of spatial econometrics in discrete or count data are considerably less established. (Glaser, n.d.; Lambert et al., 2010; Suhartono et al., 2012). In order to close this gap, a Spatial Autoregressive–Multivariate Adaptive Generalized Poisson Regression Spline (SAR–MAGPRS) estimator is suggested. Although the estimator is designed under the assumption of a generalized Poisson distribution, it can be applied to other count distributions as well. This study provides an information-theoretic framework that presents an application of the Penalized Least Square (PLS) method to estimate the SAR–MAGPRS model.

2. Materials and Methods

2.1 MAGPRS

Two adaptations to the original MARS (Friedman, 1991) that help to relax the Gaussian assumption are 'Marge' (Stoklosa & Warton, 2018) and 'Earth' (Milborrow, 2021). The R package 'marge' extends MARS to handle responses from well-known exponential families and to manage clusters of correlated data. The 'earth' package enables the fitting of models to non-normal responses by a generalized linear model over a number of basis functions. MAGPRS is a study that combines MARS and Generalized Poisson Regression (Hidayati, 2019; Hidayati et al., 2019; Otok et al., 2019; Yasmirullah, Otok, Trijoyo Purnomo, et al., 2021). The general model of MAGPRS:

$$\mu_i = \exp\left(a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} \left[s_{km} \left(x_{v(k,m)\mathrm{i}} - t_{km} \right) \right]_+ \right) \tag{1}$$

where,

a_0 : constant basis function

a_m : m – basis function

M : maximum basis function

Km : maximum interaction of m-th basis function

s_{km} : sign of basis function

$$\text{if } s_{km} = \begin{cases} +1 \text{ then } +\left(x_{v(k,m)i} - t_{km} \right)_+ & \text{where } x_{v(k,m)i} > t_{km} \\ -1 \text{ then } -\left(x_{v(k,m)i} - t_{km} \right)_+ & \text{where } x_{v(k,m)i} < t_{km} \end{cases}$$

$x_{v(k,m)i}$: predictor variable

t_{km} : knot value

2.2 SAR-MAGPRS

The new model of SAR-MAGPRS, which combines the spatial, MARS, and Generalized Poisson models, is the subject of the theoretical framework we developed in this research. The SAR-MAGPRS model in general:

$$\begin{aligned} \ln \mu_i &= f\left(\mathbf{x}_i\right) = \rho \mathbf{W} y_i + a_0 + \sum_{m=1}^{M} a_m B_{mi}\left(\mathbf{x}_i\right), \\ \mu_i &= \exp\left(\rho \mathbf{W} y_i + a_0 + \sum_{m=1}^{M} a_m B_{mi}\left(\mathbf{x}_i\right) \right), \end{aligned} \tag{2}$$

where,

$$B_{mi}\left(\mathbf{x}_i\right) = \prod_{k=1}^{K_m} \left[s_{km} \left(x_{v(k,m)\mathrm{i}} - t_{km} \right) \right]_+ \qquad \mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{1n} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & w_{n3} & \cdots & w_{nn} \end{bmatrix}$$

$B_{mi}\left(\mathbf{x}_i\right)$: basis function

ρ : spatial lag parameter for the response variable
$\mathbf{W}$: weighted matrix
y_i : response variable

2.3 PLS

Estimating the regression curve is the challenge with nonparametric regression. Minimizing the penalized least squares function is one method for estimating the regression curve (Wahba, 1990). The first step is to construct the penalized least square function:

$$\psi_i = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \mu_i\right)^2 + \eta\int_a^b \left(\mu_i^{(2)}\right)^2 dx \tag{3}$$

where,

$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \mu_i\right)^2$: ASR = Average Square Residual

$\eta\int_a^b \left(f^{(m)}\left(x_i\right)\right)^2 dx$: RP = Rougness Penalty

Then, find the first and second derivative of μ_i in equation (2) with respect to $\mathbf{x}_i$.

$$\mu_i^{(1)} = \sum_{m=1}^{M} a_m B_{mi}'\left(\mathbf{x}_i\right)\exp\left(\rho\mathbf{W}y_i + a_0 + \sum_{m=1}^{M} a_m B_{mi}\left(\mathbf{x}_i\right)\right) \tag{4}$$

$$\begin{aligned}\mu_i^{(2)} = &\sum_{m=1}^{M} a_m B_{mi}''\left(\mathbf{x}_i\right)\exp\left(\rho\mathbf{W}y_i + a_0 + \sum_{m=1}^{M} a_m B_{mi}\left(\mathbf{x}_i\right)\right)\\ &+\sum_{m=1}^{M} a_m B_{mi}'\left(\mathbf{x}_i\right)\sum_{m=1}^{M} a_m B_{mi}'\left(\mathbf{x}_i\right)\exp\left(\rho\mathbf{W}y_i + a_0 + \sum_{m=1}^{M} a_m B_{mi}\left(\mathbf{x}_i\right)\right)\end{aligned} \tag{5}$$

According to Eq. (4) and Eq. (5), the ψ_i function for the SAR-MAGPRS model is:

$$\begin{aligned}\psi_i = &\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \exp\left(\rho\mathbf{W}y_i + a_0 + \sum_{m=1}^{M} a_m B_{mi}\left(\mathbf{x}_i\right)\right)\right)^2\\ &+\eta\int_a^b\left(\sum_{m=1}^{M} a_m B_{mi}''\left(\mathbf{x}_i\right)\exp\left(\rho\mathbf{W}y_i + a_0 + \sum_{m=1}^{M} a_m B_{mi}\left(\mathbf{x}_i\right)\right)\right.\\ &\left.+\sum_{m=1}^{M} a_m B_{mi}'\left(\mathbf{x}_i\right)\sum_{m=1}^{M} a_m B_{mi}'\left(\mathbf{x}_i\right)\exp\left(\rho\mathbf{W}y_i + a_0 + \sum_{m=1}^{M} a_m B_{mi}\left(\mathbf{x}_i\right)\right)\right)^2 dx\end{aligned} \tag{6}$$

The next step is to determine the smoothing parameter, then estimate the basis function coefficient, $\left(\hat{\mathbf{a}}\right)$, and the spatial lag, $\left(\hat{\rho}\right)$.

3. Results

3.1 Parameter Estimation

The SAR-MAGPRS model can be estimated using the PLS function in Eq. (6). If the smoothing parameter is zero, then Eq. (6) becomes:

$$\begin{aligned}\boldsymbol{\psi}_{SAR\eta=0} &= \frac{1}{n}\left[\mathbf{y} - \exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right]^2\\ &= \frac{1}{n}\left[\left(\mathbf{y} - \exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\mathbf{y} - \exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)\right]\\ &= \frac{1}{n}\begin{bmatrix}\mathbf{y}'\mathbf{y} - \left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} - \left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y}\\ +\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)\end{bmatrix}\\ &= \frac{1}{n}\begin{bmatrix}\mathbf{y}'\mathbf{y} - 2\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y}\\ +\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)\end{bmatrix}\end{aligned} \tag{7}$$

After obtaining the function, the next step is slightly easier to find the estimated parameters of the model.

Theorem 1. *Suppose the MARS model can support the count data type and incorporate spatial lag autocorrelation, where the response variable is assumed to have a Generalized Poisson distribution, then the model is SAR-MAGPRS. Also, the estimated coefficient parameter of the basis function for SAR-MAGPRS,* $\hat{\mathbf{a}}$ *, is:*

$$\hat{\mathbf{a}} = \left[\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\hat{\mathbf{a}}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\hat{\mathbf{a}}\right)\right)\right]^{-1}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\hat{\mathbf{a}}\right)\right)'\mathbf{y} \tag{8}$$

Proof of Theorem 1. The estimated coefficient parameter of the basis function for SAR-MAGPRS can be found by the penalized least squares function in equation (7).

$$\begin{aligned}
\frac{\partial\left(\boldsymbol{\psi}_{SAR\eta=0}\right)}{\partial(\mathbf{a})} &= \frac{\partial\left(\frac{1}{n}\begin{bmatrix}\mathbf{y}'\mathbf{y} - 2\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} \\ +\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)\end{bmatrix}\right)}{\partial(\mathbf{a})} \\
0 &= \frac{1}{n}\begin{bmatrix}-2\mathbf{B}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} \\ +2\mathbf{B}\mathbf{a}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)\end{bmatrix} \\
\frac{1}{n}\left[2\mathbf{B}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y}\right] &= \frac{1}{n}\left[2\mathbf{B}\mathbf{a}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)\right] \\
2\mathbf{B}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} &= 2\mathbf{B}\mathbf{a}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right) \\
\mathbf{B}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} &= \mathbf{B}\mathbf{a}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right) \\
\mathbf{B}^{-1}\mathbf{B}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} &= \mathbf{B}^{-1}\mathbf{B}\mathbf{a}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right) \\
\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} &= \mathbf{a}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right) \\
\hat{\mathbf{a}} &= \left[\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\hat{\mathbf{a}}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\hat{\mathbf{a}}\right)\right)\right]^{-1}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\hat{\mathbf{a}}\right)\right)'\mathbf{y}
\end{aligned} \tag{9}$$

Theorem 2. *Suppose the MARS model can support the count data type and incorporate spatial lag autocorrelation, where the response variable is assumed to have a Generalized Poisson distribution, then the model is SAR-MAGPRS. Also, the estimated spatial lag parameter of the response variable for SAR-MAGPRS,* $(\hat{\rho})$ *, is:*

$$\hat{\rho} = \left[\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)\right]^{-1}\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} \tag{10}$$

Proof of Theorem 2. The estimated spatial lag parameter of the response variable for SAR-MAGPRS can be found by the penalized least squares function in Eq. (7).

$$\begin{aligned}
\frac{\partial\left(\boldsymbol{\psi}_{SAR\eta=0}\right)}{\partial(\rho)} &= \frac{\partial\left(\frac{1}{n}\begin{bmatrix}\mathbf{y}'\mathbf{y} - 2\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} \\ +\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)\end{bmatrix}\right)}{\partial(\rho)} \\
0 &= \frac{1}{n}\begin{bmatrix}-2\mathbf{W}\mathbf{y}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} \\ +2\rho\mathbf{W}\mathbf{y}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)\end{bmatrix} \\
\frac{1}{n}\left[2\mathbf{W}\mathbf{y}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y}\right] &= \frac{1}{n}\left[2\rho\mathbf{W}\mathbf{y}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)\right] \\
2\mathbf{W}\mathbf{y}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} &= 2\rho\mathbf{W}\mathbf{y}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right) \\
\mathbf{W}\mathbf{y}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} &= \rho\mathbf{W}\mathbf{y}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right) \\
\left(\mathbf{W}\mathbf{y}\right)^{-1}\mathbf{W}\mathbf{y}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} &= \rho\left(\mathbf{W}\mathbf{y}\right)^{-1}\mathbf{W}\mathbf{y}\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right) \\
\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y} &= \rho\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right) \\
\hat{\rho} &= \left[\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)\right]^{-1}\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y} + \mathbf{B}\mathbf{a}\right)\right)'\mathbf{y}
\end{aligned} \tag{11}$$

3.2 GCV for SAR-MAGPRS

Generalized cross-validation (GCV) is utilized in the MARS algorithm to select the optimal basis function (Friedman, 1991). This GCV formula is expressed in the following equation:

$$GCV_{SAR-MAGPRS} = \frac{MSE}{\left(1-\frac{\tilde{C}(M)}{n}\right)^2} = \left[\frac{n^{-1}\sum_{i=1}^{n}\left(y_i-\hat{f}\left(\mathbf{x_i}\right)\right)^2}{\left(1-\frac{\tilde{C}(M)}{n}\right)^2}\right] \tag{12}$$

where,

$$\tilde{C}(M) = C\left(M\right) + dM \tag{13}$$

n	:	number of data
y_i	:	response variable or actual value of data
$\hat{f}\left(\mathbf{x_i}\right)$	:	predicted value for data
$\tilde{C}\left(M\right)$	:	complex function
$C\left(M\right)$	:	number of constant and non-constant basis functions
M	:	number of non-constant basis functions
d	:	degree of interaction

The GCV formula for the SAR-MAGPRS model has $\hat{f}\left(\mathbf{x}_i\right)$ in the equation below.

$$\begin{aligned}
\hat{f}\left(\mathbf{x}_i\right) &= \ln\mu_i = \hat{\rho}\mathbf{W}y_i + \hat{a}_0 + \sum_{m=1}^{M}\hat{a}_m B_{mi}\left(\mathbf{x}_i\right) \\
&= \hat{\rho}\mathbf{W}\mathbf{y} + \mathbf{B}\hat{\mathbf{a}} \\
&= \left\{\left[\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y}+\mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y}+\mathbf{B}\mathbf{a}\right)\right)\right]^{-1}\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y}+\mathbf{B}\mathbf{a}\right)\right)'\mathbf{y}\right\}\mathbf{W}\mathbf{y} + \\
&\quad \mathbf{B}\left\{\left[\left(\exp\left(\rho\mathbf{W}\mathbf{y}+\mathbf{B}\hat{\mathbf{a}}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y}+\mathbf{B}\hat{\mathbf{a}}\right)\right)\right]^{-1}\left(\exp\left(\rho\mathbf{W}\mathbf{y}+\mathbf{B}\hat{\mathbf{a}}\right)\right)'\mathbf{y}\right\} \\
&= \left\{\left[\left(\exp\left(\rho\mathbf{W}\mathbf{y}+\mathbf{B}\hat{\mathbf{a}}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y}+\mathbf{B}\hat{\mathbf{a}}\right)\right)\right]^{-1}\left(\exp\left(\rho\mathbf{W}\mathbf{y}+\mathbf{B}\hat{\mathbf{a}}\right)\right)'\mathbf{y}\right\}\left\{\mathbf{W}\mathbf{y}+\mathbf{B}\right\}
\end{aligned} \tag{14}$$

If Eq. (14) substitutes for Eq. (12), then we have found the equation below.

$$GCV_{SAR-MAGPRS} = \left[\frac{n^{-1}\left(\begin{array}{l}\mathbf{y}-\left\{\left[\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y}+\mathbf{B}\mathbf{a}\right)\right)'\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y}+\mathbf{B}\mathbf{a}\right)\right)\right]^{-1}\left(\exp\left(\hat{\rho}\mathbf{W}\mathbf{y}+\mathbf{B}\mathbf{a}\right)\right)'\mathbf{y}\right\}\mathbf{W}\mathbf{y}+ \\ \mathbf{B}\left\{\left[\left(\exp\left(\rho\mathbf{W}\mathbf{y}+\mathbf{B}\hat{\mathbf{a}}\right)\right)'\left(\exp\left(\rho\mathbf{W}\mathbf{y}+\mathbf{B}\hat{\mathbf{a}}\right)\right)\right]^{-1}\left(\exp\left(\rho\mathbf{W}\mathbf{y}+\mathbf{B}\hat{\mathbf{a}}\right)\right)'\mathbf{y}\right\}\end{array}\right)^2}{\left(1-\frac{\tilde{C}(M)}{n}\right)^2}\right] \tag{15}$$

3.3 Application of the SAR-MAGPRS Model

Using the SAR-MAGPRS model, the number of tuberculosis cases in Lamongan, Indonesia, was analyzed. This analysis utilizes secondary data from the 2017 Lamongan health profile. Table 1 presents the research variables.

Table 1
Research Variables

Notation	Research Variables
Y	The number of tuberculosis
X_1	Population density (people/km^2)
X_2	HIV/AIDS prevalence (per 10,000 population)
X_3	Percentage of households with PHBS (%)
X_4	Percentage of healthy house (%)
X_5	Ratio of primary health facilities (per 10,000 population)
X_6	Ratio of health workers (per 10,000 population)
X_7	Percentage of the population enrolled in school (%)

The data visualization has been shown in the figure below.

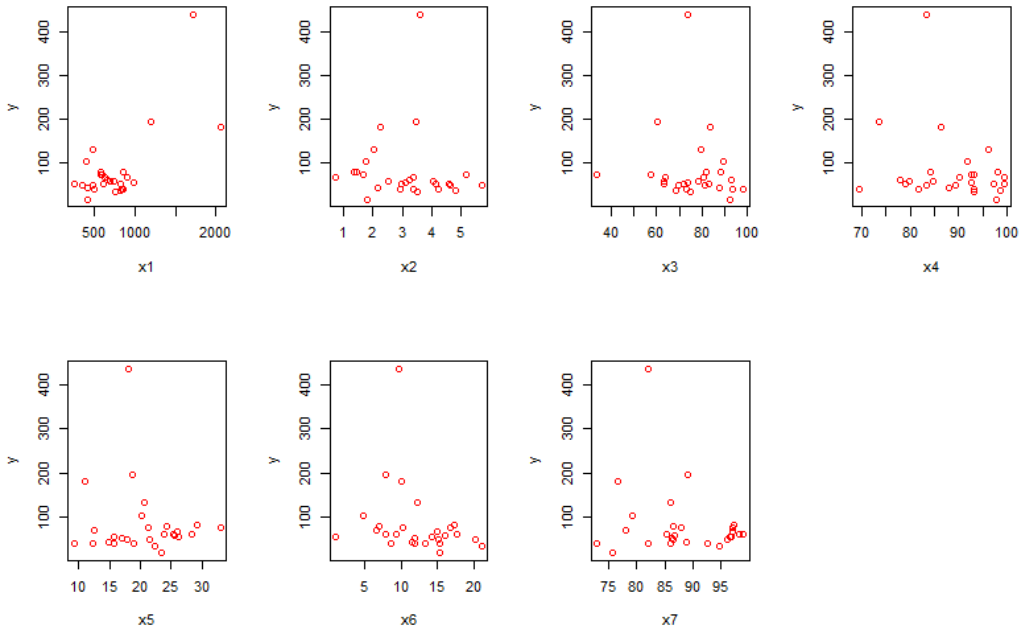

Fig. 1 Visualization of Research Variables

Fig. 1 illustrates the nonlinear relationship that exists between response and predictor variables; hence, the pattern of the plot appears random and does not follow a particular pattern. Therefore, nonparametric regression is the appropriate method for this study. This study compares two nonparametric regression approaches: MAGPRS and SAR-MAGPRS. Also, compare with the generalized linear model (GLM). The AIC value for the modeling results has been shown in Table 2.

Table 2
Comparison of GLM, MAGPRS, and SAR-MAGPRS

Model	AIC
GLM	607.36
MAGPRS	475
SAR-MAGPRS	**294.2**

The best model according to the AIC criterion is the one with the lowest AIC value. Therefore, the best model is the SAR-MAGPRS model with an AIC value of 294.2.

The best model of SAR-MAGPRS for the number of tuberculosis cases in Lamongan has a GCV value of 3325.457, where the model is:

$$\boldsymbol{\mu} = \exp\left(-0.1461\mathbf{W}\mathbf{y} + 78.4905 + 0.2239BF_1 - 8.9728BF_2\right)$$

where,

$$BF_1 = h(x_1\text{-}738.94)$$
$$BF_2 = h(18.98 - x_5)$$

Next, we will interpret one of the basis functions from the model, which is $BF_1 = h(x_1\text{-}738.94)$. This indicates that if the population density exceeds 738.94, the BF_1 coefficient will have a considerable impact. In addition, if BF_1 increases by one unit and all other basis functions remain unchanged, the number of tuberculosis cases will increase by 0.2239.

4. Conclusions

We have developed a spatial and non-parametric regression model, which is the SAR-MAGPRS model. We have demonstrated how to estimate the SAR-MAGPRS model parameter. The same method for constructing the SAR-MAGPRS model can easily be used to the other MARS models based on other distributions. The parameter estimation developed in this study using penalized least squares (PLS) methods is then necessary for the development of other estimation methods for further studies.

Acknowledgments

We would like to thank the anonymous reviewers for their informative remarks and ideas, which assisted in advancing our research.

References

Ampulembang, A. P., Otok, B. W., Rumiati, A. T., & Budiasih. (2015). Bi-responses nonparametric regression model using MARS and its properties. *Applied Mathematical Sciences*, *9*(29–32), 1417–1427. https://doi.org/10.12988/ams.2015.5127

Dey, P., & Das, A. K. (2016). Application of Multivariate Adaptive Regression Spline-Assisted Objective Function on Optimization of Heat Transfer Rate Around a Cylinder. *Nuclear Engineering and Technology*, *48*(6), 1315–1320. https://doi.org/10.1016/j.net.2016.06.011

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines (MARS). *The Annals of Statistics*, *19*(1), 1–67.

Glaser, S. (n.d.). *A Review Of Spatial Econometric Models For Count Data*. https://wiso.uni-hohenheim.de/papers

Hidayati, S. (2019). *Penaksiran Parameter dan Statistik Uji Model Multivariate Adaptive Generalized Poisson Regression Spline pada Kasus Jumlah Penderita Ispa pada Bayi di Surabaya Tahun 2017* [Tesis]. Institut Teknologi Sepuluh Nopember.

Hidayati, S., Otok, B. W., & Purhadi. (2019). Parameter Estimation and Statistical Test in Multivariate Adaptive Generalized Poisson Regression Splines. *IOP Conference Series: Materials Science and Engineering*, *546*(5). https://doi.org/10.1088/1757-899X/546/5/052051

Kooperberg, C. (2014). *Multivariate Adaptive Regression Splines*. Wiley StatsRef: Statistics Reference Online.

Lambert, D. M., Brown, J. P., & Florax, R. J. G. M. (2010). A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application. *Regional Science and Urban Economics*, *40*(4), 241–252. https://doi.org/10.1016/j.regsciurbeco.2010.04.001

Liu, L., Zhang, S., & Cheng, Y. M. (2019). Advanced reliability analysis of slopes in spatially variable soils using multivariate adaptive regression splines. *Geoscience Frontiers*, *10*(2), 1–12. https://doi.org/10.1016/j.gsf.2018.03.013

Makalew, L. A., Kuntoro, Otok, B. W., Soenarnatalina, M., & Layuk, S. (2019). Modeling the number of cases of tuberculosis sensitive drugs (Tbsd) in East Java using geographically weighted poisson regression (GWPR). *Indian Journal of Public Health Research and Development*, *10*(6), 398–403. https://doi.org/10.5958/0976-5506.2019.01305.6

Makalew, L. A., Otok, B. W., & Barung, E. N. (2019). Spatio of lungs Tuberculosis (Tb Lungs) in East Java Using Geographically Weighted Poisson Regression (GWPR). In *Indian Journal of Public Health Research & Development* (Vol. 10, Issue 8).

Milborrow, S. (2021). Package Earth. In *The Annals of Statistics* (Vol. 19, Issue 1). https://cran.r-project.org/web/packages/earth/earth.pdf

Otok, B. W., Hidayati, S., & Purhadi. (2019). Multivariate Adaptive Generalized Poisson Regression Spline (MAGPRS) on the number of acute respiratory infection infants. *Journal of Physics: Conference Series*, *1397*(1). https://doi.org/10.1088/1742-6596/1397/1/012062

Otok, B. W., Putra, R. Y., & P Yasmirullah, S. D. (2020). Bootstrap Aggregating Multivariate Adaptive Regression Spline For Observational Studies In Diabetes Cases. In *Systematic Reviews in Pharmacy* (Vol. 11, Issue 8).

Raj, N., & Gharineiat, Z. (2021). Evaluation of multivariate adaptive regression splines and artificial neural network for prediction of mean sea level trend around northern australian coastlines. *Mathematics*, *9*(21). https://doi.org/10.3390/math9212696

Stoklosa, J., & Warton, D. I. (2018). A Generalized Estimating Equation Approach to Multivariate Adaptive Regression Splines. *Journal of Computational and Graphical Statistics*, *27*(1), 245–253. https://doi.org/10.1080/10618600.2017.1360780

Suhartono, Faulina, R., Lusia, D. A., Otok, B. W., Sutikno, & Kuswanto, H. (2012). Ensemble method based on ANFIS-ARIMA for rainfall prediction. *ICSSBE 2012 - Proceedings, 2012 International Conference on Statistics in Science, Business and Engineering: "Empowering Decision Making with Statistical Sciences,"* 240–243. https://doi.org/10.1109/ICSSBE.2012.6396564

Wang, X., Yang, C., & Zhou, M. (2021). Partial least squares improved multivariate adaptive regression splines for visible and near-infrared-based soil organic matter estimation considering spatial heterogeneity. *Applied Sciences (Switzerland)*, *11*(2), 1–16. https://doi.org/10.3390/app11020566

Yasmirullah, S. D. P., Otok, B. W., Purnlmo, J. D. T., & Prastyo, D. D. (2021). Multivariate adaptive regression spline (MARS) methods with application to multi drug-resistant tuberculosis (mdr-tb) prevalence. *AIP Conference Proceedings*, *2329*. https://doi.org/10.1063/5.0042145

Yasmirullah, S. D. P., Otok, B. W., Purnomo, J. D. T., & Prastyo, D. D. (2021). Parameter Estimation of Multivariate Adaptive Regression Spline (MARS) with Stepwise Approach to Multi Drug-Resistant Tuberculosis (MDR-TB) Modeling in Lamongan Regency. *Journal of Physics: Conference Series*, *1752*(1). https://doi.org/10.1088/1742-6596/1752/1/012017

Yasmirullah, S. D. P., Otok, B. W., Trijoyo Purnomo, J. D., & Prastyo, D. D. (2021). Modification of Multivariate Adaptive Regression Spline (MARS). *Journal of Physics: Conference Series*, *1863*(1). https://doi.org/10.1088/1742-6596/1863/1/012078

York, T. P., Eaves, L. J., & van den Oord, E. J. C. G. (2006). Multivariate adaptive regression splines: A powerful method for detecting disease-risk relationship differences among subgroups. *Statistics in Medicine*, *25*(8), 1355–1367. https://doi.org/10.1002/sim.2292

Zheng, G., Yang, P., Zhou, H., Zeng, C., Yang, X., He, X., & Yu, X. (2019). Evaluation of the earthquake induced uplift displacement of tunnels using multivariate adaptive regression splines. *Computers and Geotechnics*, *113*. https://doi.org/10.1016/j.compgeo.2019.103099