Geographically Weighted Negative Binomial Regression—incorporating overdispersion

**Alan Ricardo da Silva ·
Thais Carvalho Valadares Rodrigues**

Abstract Global regression assumes that a single model adequately describes all parts of a study region. However, the heterogeneity in the data may be sufficiently strong that relationships between variables can not be spatially constant. In addition, the factors involved are often sufficiently complex that it is difficult to identify them in the form of explanatory variables. As a result Geographically Weighted Regression (GWR) was introduced as a tool for the modeling of non-stationary spatial data. Using kernel functions, the GWR methodology allows the model parameters to vary spatially and produces non-parametric surfaces of their estimates. To model count data with overdispersion, it is more appropriate to use a negative binomial distribution instead of a Poisson distribution. Therefore, we propose the Geographically Weighted Negative Binomial Regression (GWNBR) method for the modeling of data with overdispersion. The results obtained using simulated and real data show the superiority of this method for the modeling of non-stationary count data with overdispersion compared with competing models, such as global regressions, e.g., Poisson and negative binomial and Geographically Weighted Poisson Regression (GWPR). Moreover, we illustrate that these competing models are special cases of the more robust model GWNBR.

Keywords Generalized linear models · Poisson distribution · Negative binomial distribution · Overdispersion · Spatial analysis · Geographically Weighted Regression

A.R. da Silva (✉) · T.C.V. Rodrigues
Departamento de Estatistica, Universidade de Brasilia, Brasilia, Brazil
e-mail: alansilva@unb.br

T.C.V. Rodrigues
e-mail: tatahcv@yahoo.com.br

1 Introduction

Geographically Weighted Regression (GWR) allows the spatial modeling of non-stationary data and was defined by Fotheringham et al. (2002). However, this technique is used when the distribution of the data is Gaussian. In many applications the dependent variable represents a count which makes classic GWR inappropriate to model this type of data. Poisson and negative binomial are the most adequate distributions for the modeling of count data, and Geographically Weighted Poisson Regression (GWPR) was developed by Nakaya et al. (2005). Kobayashi and Lane (2007) and Hadayeghi et al. (2010) used GWPR because Geographically Weighted Regression for a negative binomial distribution was not available. The advantage of a negative binomial distribution is the ability to model data with overdispersion (Hilbe 2011) because this type of data has an additional parameter, α. Furthermore, this distribution is a generalization of the geometric and Poisson distributions, when $\alpha = 1$ and $\alpha \to 0$, respectively.

The negative binomial and Poisson distributions belong to the exponential family of distributions (Nelder and Wedderburn 1972). The Fisher Score algorithm of Generalized Linear Models (GLM) can be extended to the spatial case with a negative binomial distribution considering a modification in the Iteratively Reweighted Least Squares (IRLS) procedure, which results in the proposed Geographically Weighted Poisson Regression (GWPR). The parameter α of the negative binomial, which is supposedly known through the IRLS method, can be estimated in a subroutine with the maximum likelihood method using the Newton-Raphson (NR) algorithm (Hilbe 2011).

Since the work of Nakaya et al. (2005), many other studies have attempted to investigate the limitations and problems associated with Geographically Weighted Regression, such as spatial error autocorrelation (Leung et al.

2000b), multicollinearity (Wheeler and Tiefelsdorf 2005; Geniaux et al. 2011) and extreme coefficients including sign reversal (Farber and Paez 2007). In a more recent work, Paez et al. (2011) investigated spatially varying relationships in a simulation-based study using diverse levels of multicollinearity between the explicative variables. These researches concluded that caution should be exercised when attempting to interpret coefficients obtained through Gaussian GWR.

The objective of this paper is to extend the IRLS and NR methods to establish Geographically Weighted Regression for a negative binomial distribution, which is denoted GWNBR, for the modeling of non-stationary count spatial data with overdispersion. In addition, the procedure developed by Paez et al. (2011) will be used to investigate the performance of our proposed method in the presence of multicollinearity between the explicative variables. Section 2 presents the specifications of the GWNBR methodology. Section 3 presents some simulations and real applications that were used to evaluate the potential of GWNBR compared with GWPR and to conduct the simulations presented by Paez et al. (2011). Some conclusions are drawn in Sect. 4.

2 Specifications of Geographically Weighted Negative Binomial Regression

The most used global Negative Binomial Regression (NB-2) considers a logarithm link function. Parameterizing this model in terms of μ_j / t_j, where t_j is an offset variable, μ_j is the predicted mean, α is the parameter of overdispersion, β_k is the parameter related to the explicative variable x_k, for $k = 1, \ldots, K$, and y_j is the j-th dependent variable for $j = 1, \ldots, n$ we have

$$y_j \sim \text{NB}\left[t_j \exp\left(\sum_k \beta_k x_{jk}\right), \alpha\right] \tag{1}$$

where NB represents Negative Binomial.

GWNBR is an extension of the global or non-spatial model (1), which allows the spatial variation of the parameters β_k and α. Without limiting the functional form of this variation, GWNBR produces non-parametric surfaces of the parameter estimates. This local model is described as the following:

$$y_j \sim \text{NB}\left[t_j \exp\left(\sum_k \beta_k(u_j, v_j) x_{jk}\right), \alpha(u_j, v_j)\right] \tag{2}$$

where (u_j, v_j) are the locations (coordinates) of the data points j, for $j = 1, \ldots, n$.

The parameter estimation of the global model (1) is performed interactively with the combination of the NR and IRLS methods, i. e., using $\hat{\alpha}$, which is estimated by the NR

method, we estimate the vector $\boldsymbol{\beta}$ using the IRLS method. Thus, from this new $\hat{\boldsymbol{\beta}}$, we update $\hat{\alpha}$, and so forth until convergence is obtained. However, modifications in the NR and IRLS algorithms are necessary to incorporate the local variations.

Let us consider first the modified IRLS method. Initially, suppose that the parameters $\boldsymbol{\alpha}(u, v)$ are known. Then, the log-likelihood of GWNBR, as function of $\boldsymbol{\beta}(u, v)$, is given by

$$
L\big(\boldsymbol{\beta}(u_j, v_j) | x_{jk}, y_j, \alpha_j\big)
$$
$$
= \sum_{j=1}^{n} \big\{ y_j \log\left(\alpha_j \mu_j\right) - (y_j + 1/\alpha_j)
$$
$$
\times \log\left(1 + \alpha_j \mu_j\right) + \log\left[\Gamma(y_j + 1/\alpha_j)\right]
$$
$$
- \log\left[\Gamma(1/\alpha_j)\right] - \log\left[\Gamma(y_j + 1)\right] \big\} \tag{3}
$$

where for $j = 1, \ldots, n$,

$$\mu_j = t_j \exp\left(\sum_k \beta_k(u_j, v_j) x_{jk}\right) \tag{4}$$

$$\alpha_j = \alpha(u_j, v_j) \tag{5}$$

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \tag{6}$$

Considering that the surfaces of the parameters $\boldsymbol{\beta}(u, v)$ are approximately planar (Fig. 1) in the neighborhood of the regression point i, the local log-likelihood of GWNBR at location i can be approximated as

$$
L\big(\boldsymbol{\beta}(u_i, v_i) | x_{jk}, y_j, \alpha_i\big)
$$
$$
= \sum_{j=1}^{n} \big\{ y_j \log\left[\alpha_i \mu_j\big(\boldsymbol{\beta}(u_i, v_i)\big)\right]
$$
$$
- [y_j + 1/\alpha_i] \log\left[1 + \alpha_i \mu_j\big(\boldsymbol{\beta}(u_i, v_i)\big)\right]
$$
$$
+ \log\left[\Gamma(y_j + 1/\alpha_i)\right] - \log\left[\Gamma(1/\alpha_i)\right]
$$
$$
- \log\left[\Gamma(y_j + 1)\right] \big\} w(d_{ij}) \tag{7}
$$

where, for $i = 1, \ldots, g$, $\mu_j(\boldsymbol{\beta}(u_i, v_i))$ is the predicted mean at location j with the parameters at regression point i:

$$\mu_j\big(\boldsymbol{\beta}(u_i, v_i)\big) = t_j \exp\left(\sum_k \beta_k(u_i, v_i) x_{jk}\right) \tag{8}$$

and $w(d_{ij})$ is the geographical weight of observation j at regression point i, which depends on the distance between them. A kernel function can be used to determine these weights, as will be described in Sect. 3.

Note that it is possible to estimate the parameters at any regression point i. However, the estimated means are calculated only for the observed points j at which the information x_{jk} is known.
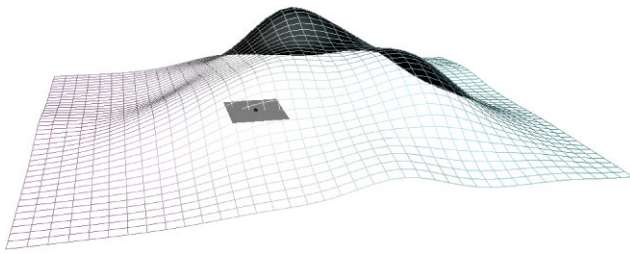
Fig. 1 Surface of β_k which is locally planar at point i

The maximum local log-likelihood (7) provides estimates for the parameters $\boldsymbol{\beta}(u_i, v_i)$ of GWNBR at location i. Thus, using the results provided by Nakaya et al. (2005) for GWPR, the analytical solution for (7) is given by

$$\hat{\boldsymbol{\beta}}(u_i, v_i)^{(m+1)} = \left[\mathbf{X}'\mathbf{W}(u_i, v_i)\mathbf{A}(u_i, v_i)^{(m)}\mathbf{X}\right]^{-1}\mathbf{X}'$$
$$\times \mathbf{W}(u_i, v_i)\mathbf{A}(u_i, v_i)^{(m)}\mathbf{z}(u_i, v_i)^{(m)} \quad (9)$$

where $\mathbf{X}$ is an $n \times k$ matrix of the explicative variables

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix} \quad (10)$$

$\mathbf{W}(u_i, v_i)$ is an $n \times n$ GWR diagonal weighting matrix for point i

$$\mathbf{W}(u_i, v_i) = \begin{pmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{in} \end{pmatrix} \quad (11)$$

and $\mathbf{A}(u_i, v_i)^{(m)}$ is an $n \times n$ GLM diagonal weighting matrix for iteration m and location i

$$\mathbf{A}(u_i, v_i)^{(m)} = \begin{pmatrix} a_{i1}^{(m)} & 0 & \dots & 0 \\ 0 & a_{i2}^{(m)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{in}^{(m)} \end{pmatrix} \quad (12)$$

$$\mathbf{z}(u_i, v_i)^{(m)} = X\hat{\boldsymbol{\beta}}(u_i, v_i)^{(m)}$$
$$+ \frac{y_j - \mu_j(\hat{\boldsymbol{\beta}}(u_i, v_i)^{(m)})}{a_{ij}^{(m)}(1 + \alpha_i \times \mu_j(\hat{\boldsymbol{\beta}}(u_i, v_i)^{(m)}))} \quad (13)$$

In this paper, the IRLS method will be used with the observed Fisher Information Matrix. Thus, the elements $a_{ij}^{(m)}$ ($j = 1, \dots, n$) of GWNBR (12) are the following:

$$a_{ij}^{(m)} = \frac{\mu_j(\hat{\boldsymbol{\beta}}(u_i, v_i)^{(m)})}{1 + \alpha_i \mu_j(\hat{\boldsymbol{\beta}}(u_i, v_i)^{(m)})}$$
$$+ \frac{[y_j - \mu_j(\hat{\boldsymbol{\beta}}(u_i, v_i)^{(m)})][\alpha_i \mu_j(\hat{\boldsymbol{\beta}}(u_i, v_i)^{(m)})]}{1 + 2\alpha_i \mu_j(\hat{\boldsymbol{\beta}}(u_i, v_i)^{(m)}) + \alpha_i^2 \mu_j^2(\hat{\boldsymbol{\beta}}(u_i, v_i)^{(m)})} \quad (14)$$

As in GWPR, the covariance matrix of the parameter estimates can be estimated by

$$\widehat{Cov}\left[\hat{\boldsymbol{\beta}}(u_i, v_i)\right] = \mathbf{C}(u_i, v_i)\mathbf{A}(u_i, v_i)^{-1}\mathbf{C}'(u_i, v_i) \quad (15)$$

where

$$\mathbf{C}(u_i, v_i) = \left[\mathbf{X}'\mathbf{W}(u_i, v_i)\mathbf{A}(u_i, v_i)\mathbf{X}\right]^{-1}$$
$$\times \mathbf{X}'\mathbf{W}(u_i, v_i)\mathbf{A}(u_i, v_i) \quad (16)$$

and the elements of $\mathbf{W}(u_i, v_i)$ and $\mathbf{A}(u_i, v_i)$ are given by (11) and (12), respectively.

After an estimate of $\boldsymbol{\beta}(u_i, v_i)$ is obtained, the parameters α_i will be estimated using the NR method based on the local log-likelihood (7). To facilitate the calculation, the parameters r_i, where $r_i = 1/\alpha_i$, will be estimated first. Thus, rewriting (7), we have

$$L\left(r_i | y_j, \boldsymbol{\beta}(u_i, v_i)\right)$$
$$= \sum_{j=1}^{n}\{y_j \log\left[\mu_j\left(\boldsymbol{\beta}(u_i, v_i)\right)\right] - [y_j + r_i]$$
$$\times \log\left[r_i + \mu_j\left(\boldsymbol{\beta}(u_i, v_i)\right)\right] + r_i \log[r_i]$$
$$+ \log\left[\Gamma(y_j + r_i)\right] - \log\left[\Gamma(r_i)\right]$$
$$- \log\left[\Gamma(y_j + 1)\right]\}w(d_{ij}) \quad (17)$$

Maximizing the local log-likelihood (17) using the univariate NR method, we obtain

$$r_i^{(m+1)} = r_i^{(m)} - \left[H_i^{(m)}\right]^{-1}U_i^{(m)} \quad (18)$$

where $U_i^{(m)}$ and $H_i^{(m)}$ are the first and second derivatives of the local log-likelihood with respect to $r_i^{(m)}$, i.e.,

$$U_i^{(m)} = \frac{dL(r_i)}{dr_i}$$
$$= \sum_{j=1}^{n}\left\{\psi\left[r_i^{(m)} + y_j\right] - \psi\left[r_i^{(m)}\right] + \log\left[r_i^{(m)}\right]\right.$$
$$+ 1 - \log\left[r_i^{(m)} + \mu_j\left(\boldsymbol{\beta}(u_i, v_i)\right)\right]$$
$$\left. - \frac{r_i^{(m)} + y_j}{r_i^{(m)} + \mu_j\left(\boldsymbol{\beta}(u_i, v_i)\right)}\right\}w(d_{ij}) \quad (19)$$

$$
\begin{aligned}
H_i^{(m)} &= \frac{d^2 L(r_i)}{dr_i^2} \\
&= \sum_{j=1}^{n} \left\{ \psi'\left[r_i^{(m)} + y_j\right] - \psi'\left[r_i^{(m)}\right] + \frac{1}{r_i^{(m)}} \right. \\
&\quad - \frac{2}{r_i^{(m)} + \mu_j(\boldsymbol{\beta}(u_i, v_i))} \\
&\quad \left. + \frac{r_i^{(m)} + y_j}{[r_i^{(m)} + \mu_j(\boldsymbol{\beta}(u_i, v_i))]^2} \right\} w(d_{ij})
\end{aligned}
\tag{20}
$$

where $\psi(.)$ e $\psi'(.)$ are the digamma and trigamma functions, respectively, which are given by

$$
\psi(z) = \frac{d \log \Gamma(z)}{dz} \quad \text{and} \quad \psi'(z) = \frac{d\psi(z)}{dz} = \frac{d^2 \log \Gamma(z)}{dz^2}
$$

Using delta method, we find that, for a function $g(.)$ that is differentiable in θ, if the distribution $\hat{\theta}_n \to N(\theta, \sigma_n^2)$ achieves convergence, then the distribution $g(\hat{\theta}_n) \to N(g(\theta), [g'(\theta)]^2 \times \sigma_n^2)$ also converges (Casella and Berger 2001). If, under certain conditions, the estimators that maximize the local likelihood are asymptotically normal, unbiased and consistent (Staniswalis 1989), then

$$
\hat{\alpha}_i = \frac{1}{\hat{r}_i} \quad \text{and} \quad Var(\hat{\alpha}_i) = \frac{-1}{H_i \hat{r}_i^4}
\tag{21}
$$

because $Var(\hat{r}_i) = -1/H_i$, where H_i is given equation by (20), and $[g'(r_i)]^2 = 1/r_i^4$.

Thus, for each regression point i, the NR and IRLS algorithms are used alternately until the parameter estimates achieve convergence.

To complete the fitting of the model, it is necessary to estimate the bandwidth of the chosen kernel function, which helps determine the weights $w(d_{ij})$. One possibility could be to estimate it such that it minimizes the corrected AIC criterion (AICc):

$$
AIC_c = -2L(\boldsymbol{\beta}, \alpha) + 2k + \frac{2k(k+1)}{n-k-1}
\tag{22}
$$

where k is the effective number of parameters and $L(\boldsymbol{\beta}, \alpha)$ is the log-likelihood of GWNBR shown in (3). Hereafter, the term "AIC" represents AICc.

The effective number of parameters of GWNBR can be written as $k = k_1 + k_2$, where k_1 and k_2 are the effective number of parameters due to $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, respectively. Following the method developed by Nakaya et al. (2005), k_1 is given by the trace of the matrix $\mathbf{R}$, which is given by elements

$$
\boldsymbol{r}_j = \mathbf{X}_j \left[\mathbf{X}'\mathbf{W}(u_j, v_j)\mathbf{A}(u_j, v_j)\mathbf{X} \right]^{-1} \mathbf{X}'\mathbf{W}(u_j, v_j)\mathbf{A}(u_j, v_j)
\tag{23}
$$

where $\mathbf{X}_j$ is the j-th row of $\mathbf{X}$.

However, to date, it has not been possible to estimate k_2, i.e., the contribution of the surface of $\hat{\boldsymbol{\alpha}}$ on the effective number of parameters of the model. Consequently, we opted to estimate the bandwidth using the cross-validation criterion given by (Fotheringham et al. 2002):

$$
CV = \sum_{j=1}^{n} \left[y_j - \hat{y}_{\neq j}(b) \right]^2
\tag{24}
$$

where $\hat{y}_{\neq j}(b)$ is the estimated value for point j, omitting the observation j and b is the bandwidth.

Note that the indetermination of k_2 does not prevent the adjustment of GWNBR. However, this indetermination makes it difficult to compare the models because the complexity of the GWNBR, which is given by the effective number of parameters, is unknown.

2.1 GWNBRg model

To avoid the difficult associate with the estimation of k_2, we propose the Geographically Weighted Negative Binomial Regression with α global methodology, which is namely GWNBRg. In this model, the spatial variation is allowed only for $\boldsymbol{\beta}(u_i, v_i)$, i. e.,

$$
y_j \sim BN \left[t_j \exp\left(\sum_k \beta_k(u_j, v_j)x_{jk} \right), \alpha \right]
\tag{25}
$$

where the parameters are the same as in (2).

In the GWNBRg model, the estimation of the parameter α is made globally, i.e., we assume that all of the parameters in the model are stationary, and we estimate a global overdispersion $\hat{\alpha}$ to be used in the local estimates $\boldsymbol{\beta}(u_i, v_i)$. Consequently, we propose that the estimate of α in the GWNBRg model will be the same as that obtained through non-spatial (or global) negative binomial regression.

The parameters $\boldsymbol{\beta}(u_i, v_i)$ are estimated using the IRLS method, as in the GWNBR model described earlier, assuming that $\hat{\alpha}_i = \hat{\alpha}$ for all i. Note that it is not necessary to alternate the NR and IRLS methods, because once $\hat{\alpha}$ is estimated globally, we estimate $\boldsymbol{\beta}(u_i, v_i)$ for each regression point i using only the IRLS method.

Because there is no spatial variation for α, its contribution to the effective number of the parameters in the model is the unity, i.e., $k_2 = 1$. Consequently, the bandwidth can be found using the AIC criterion (22), where

$$
\begin{aligned}
&L\big(\boldsymbol{\beta}(u_i, v_i), \alpha | x_{jk}, y_j\big) \\
&= \sum_{j=1}^{n} \big\{ y_j \log(\alpha\mu_j) - (y_j + 1/\alpha) \log(1 + \alpha\mu_j) \\
&\quad + \log\left[\Gamma(y_j + 1/\alpha)\right] - \log\left[\Gamma(1/\alpha)\right] \\
&\quad - \log\left[\Gamma(y_j + 1)\right] \big\}
\end{aligned}
\tag{26}
$$

and $k = \text{tr}(\mathbf{R}) + 1$.

Thus, the objective of the GWNBRg model is to perform the GWNBR methodology in a simplified manner, which has the disadvantage of obtaining a biased estimate of α and the advantage of being a model with known complexity. We hope that the GWNBRg methodology be a better approach for the modeling of spatial count data with overdispersion than the GWPR method, which considers that α is equal to zero.

3 Simulations and applications

This section aims to show the fit of the GWNBR and GWN-BRg models to simulated and real data. Using simulated data, our objective is to evaluate whether the models correctly reproduce the parameters. The real data was used to show a practical application of the proposed model. In both cases, the results will be compared with GWPR and Negative Binomial (NBR) and Poisson (PR) regressions.

It is important to mention that the discussion presented here is mainly focused on the comparison between different distributions for the modeling of count data. The analysis of the limitations and problems associated with the GWR framework, such as the spatial error autocorrelation (Leung et al. 2000b), multicollinearity (Wheeler and Tiefelsdorf 2005; Geniaux et al. 2011) and extreme coefficients including sign reversal (Farber and Paez 2007) will not be treated as in previous studies, but some of the results show that our approach does retain some of these problems.

The simulations will be conducted using the SAS 9.2 software and the procedure presented by Paez et al. (2011), i.e., we will simulate spatial data and spatial coefficients, and we will introduce varying levels of correlations between variables x_1 and x_2 (from -0.75 to 0.75 by 0.25 and 100 times for each level), using a transformation similar to that used by Wheeler and Tiefelsdorf (2005):

$$x_2^\theta = \sin(\theta)x_1 + \cos(\theta)x_2 \qquad (27)$$

where θ is the correlation level.

The correlation between variables x_1 and x_2 and the correlation measured on simulated data are presented below. As shown, the pattern is approximately the same.

```
    Corr.
Between x1      Corr.
  and x2       Measured

  -0.75        -0.80719
  -0.50        -0.68474
  -0.25        -0.39143
   0           -0.08982
   0.25         0.27470
   0.50         0.52585
   0.75         0.75624
```
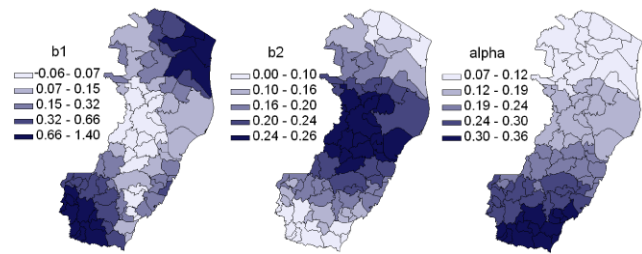

Fig. 2 True coefficients of surfaces b_1, b_2, and α of the GWNBR framework

3.1 Simulation of GWNBR

The main objective of the GWNBR and GWNBRg frameworks is the modeling of non-stationary count data with overdispersion. Thus, we first simulate data with these characteristics using the following structure:

$$y_j \sim \text{NB}\big[\mu(u_j, v_j), \alpha(u_j, v_j)\big] \qquad (28)$$

where

$$\mu(u_j, v_j) = \exp\big\{b_0(u_j, v_j) + b_1(u_j, v_j)x_{j1} + b_2(u_j, v_j)x_{j2}\big\}$$

$$b_0(u_j, v_j) = 1$$

$$b_1(u_j, v_j) = \sqrt{[u_j - \overline{u}_j]^2 \times [v_j - \overline{v}_j]^2}$$

$$b_2(u_j, v_j) = 2\big\{-\big[(u_j - \overline{u}_j)/5\big]^2 - \big[(v_j - \overline{v}_j)/5\big]^2 \qquad (29)$$
$$+ 0.13\big\}$$

$$\overline{u}_j = -40.7693, \qquad \overline{v}_j = -19.5915$$

$$\alpha(u_j, v_j) = 2\big\{10^{-7}[|v_j|]^5\big\}^2$$

The explicative variables $x_{j1}*$ and $x_{j2}*$ were initially simulated from a Uniform$(0, 1)$ distribution, and the variable x_1 was then updated to be spatially dependent by $x_{j1} := (\frac{x_{j1}*|u_j|}{1000})^2 + (\frac{x_{j1}*|v_j|}{90})^2$ and the variable x_2 was updated such that it is correlated with x_1 using (27). We use the shape of the Brazilian state Espirito Santo as a geographical reference to simulate the data (77 municipalities) because irregular tessellations are more representative of real geographic systems, as described by Farber et al. (2009). Figure 2 shows the spatial distributions of the parameters, and Fig. 3 shows the spatial distributions of the variables, assuming that the correlation between x_1 and x_2 is equal to zero.

To confirm whether the method used to find the best bandwidth affects the results based on the suggestion provided by Farber and Paez (2007), we fit the GWNBR, GWN-BRg, and GWPR using AIC and Cross-Validation (CV) approaches, but the comparison between the models will be performed using log-likelihood. To determinate the geographical weights using the nearest-neighbor bandwidth, we
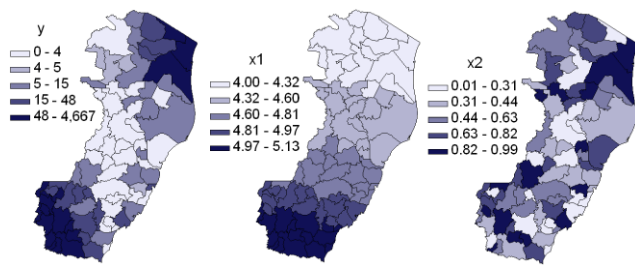
Fig. 3 Spatial patterns of y, x_1, and x_2 of the GWNBR framework

Table 1 Frequency of the best model for the fitting of GWNBR simulated data

Corr	GWNBR	GWNBRg	GWPR	NB	Total
−0.75	66	1	33	0	100
−0.50	51	3	46	0	100
−0.25	65	0	35	0	100
0	69	1	28	2	100
0.25	60	1	39	0	100
0.50	71	1	28	0	100
0.75	73	4	23	0	100
Total	455	11	232	2	700
	65.00 %	1.57 %	33.14 %	0.29 %	100 %

will use the adaptive bi-square kernel given by

$$w_{ij} = \begin{cases} [1 - (d_{ij}/b)^2]^2, & \text{if } j \text{ is one of the } N\text{-th nearest} \\ & \text{neighbors of } i; \\ 0, & \text{otherwise.} \end{cases}$$

(30)

where d_{ij} is the distance between the j-th and i-th nearest neighbors and b is the bandwidth.

To determine the geographical weights by the distance bandwidth, we will use the Gaussian kernel function given by

$$w_{ij} = \exp\left\{-\frac{1}{2}(d_{ij}/b)^2\right\}$$

(31)

where d_{ij} is the distance between the j-th and i-th locations and b is the bandwidth.

Table 1 shows the frequency of the best model using cross-validation or the AIC approach for the distance or the nearest-neighbor bandwidths. As shown, approximately 65 % of the best fits were obtained using the GWNBR method, and approximately 33 % of the best fits were obtained from the GWPR framework.

Table 2 shows the frequency of the best model using only cross-validation for the distance or the nearest-neighbor bandwidths. As shown, the percentage of times
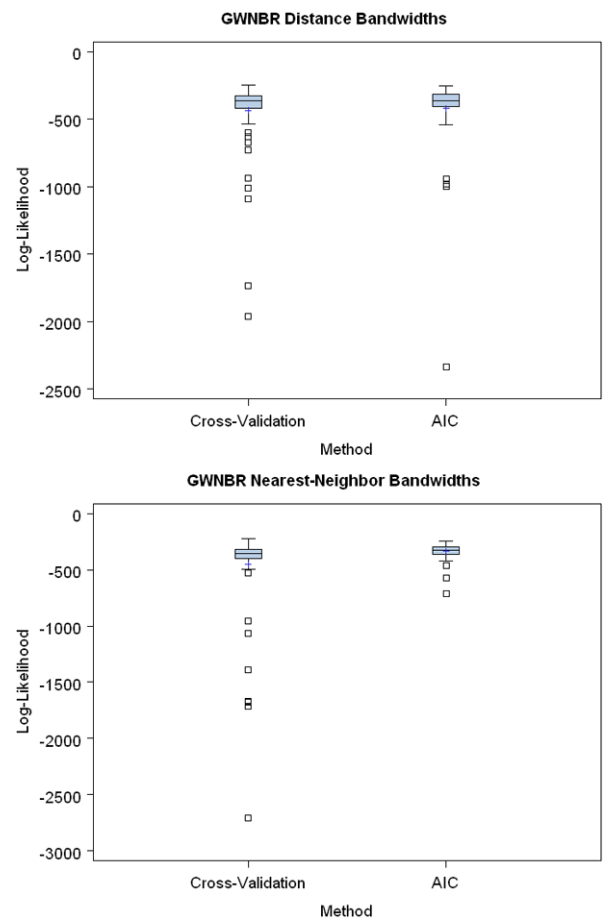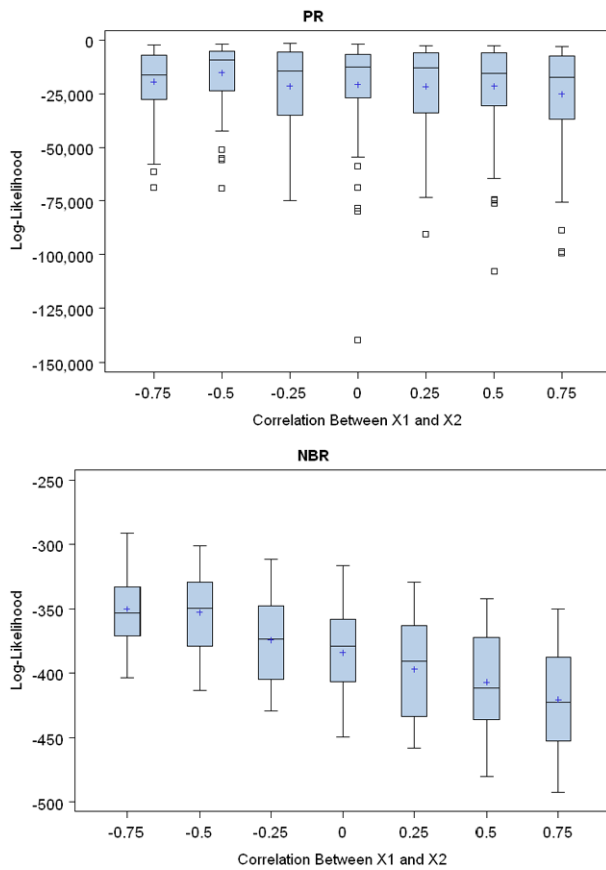

Fig. 4 Distributions of log-likelihood of GWNBR

Table 2 Frequency of the best model for the fitting of GWNBR simulated data

Corr	GWNBR	GWNBRg	GWPR	NB	Total
−0.75	79	10	5	6	100
−0.50	69	12	14	5	100
−0.25	69	12	6	13	100
0	69	14	8	9	100
0.25	68	17	9	6	100
0.50	78	17	1	4	100
0.75	77	15	5	3	100
Total	509	97	48	46	700
	72.71 %	13.86 %	6.86 %	6.57 %	100 %

that GWNBR is chosen as the best model increases to 72 %, and increases to 14 % for GWNBRg.

Figures 4 and 5 show the variability of the log-likelihoods for the GWNBR for no correlation case and PR and NBR models, respectively. In general, when the AIC was minimized to find the optimal bandwidth, the variability of the log-likelihood was smaller than when the CV was used to

Fig. 5 Distributions of log-likelihood of PR and NBR

Fig. 6 Distributions of the optimal bandwidth of GWNBR

find the optimal bandwidth. We found the same results for each level of correlation between x_1 and x_2 and for GWN-BRg and GWPR. Also, it is interesting to note that the NBR model is influenced by the correlation between variables x_1 and x_2, whereas the Poisson regression is not.

Because the cross-validation approach apparently gives the best results for the modeling of count data, let us analyze the variability in the optimal bandwidth. Figure 6 shows the box-plot of the optimal bandwidth using cross-validation or AIC for the distance and nearest-neighbor bandwidths of GWNBR. Figure 7 shows the box-plot of the optimal bandwidth using cross-validation or AIC for the distance and nearest-neighbor bandwidths of GWPR. In addition, Fig. 8 shows the box-plot of the optimal bandwidth using cross-validation or AIC and for the distance and nearest-neighbor bandwidths of GWNBRg.

In summary, the correlation between variables x_1 and x_2 appears to influence the distance bandwidth used in the GWNBR framework, but this correlation does not affect the results when the type of bandwidth is nearest-neighbor and when the model is GWPR or GWNBRg.
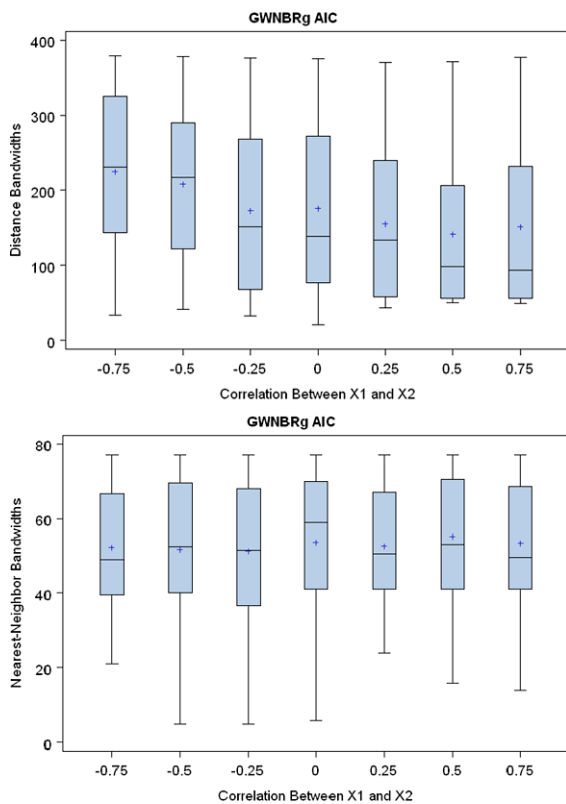
Table 3 shows some median statistics for the goodness of fit. Column b refers to the bandwidth, which is expressed in terms of the numbers of points included in the local re-
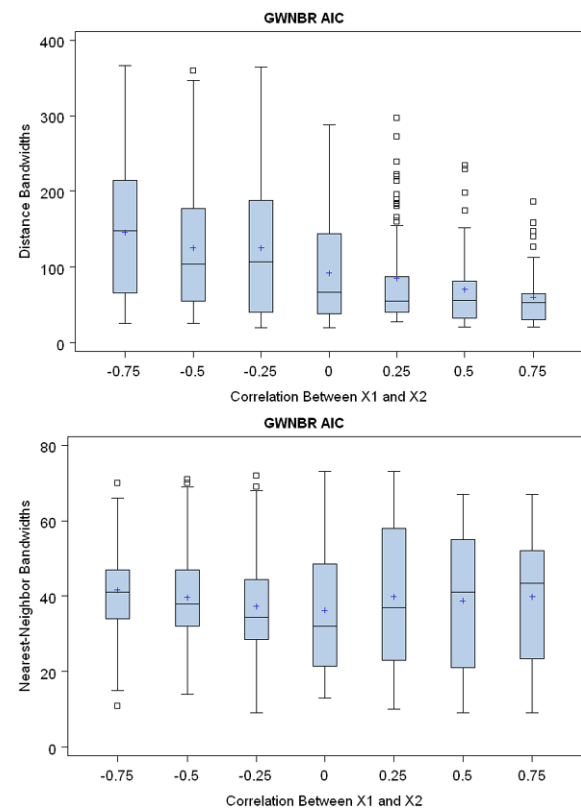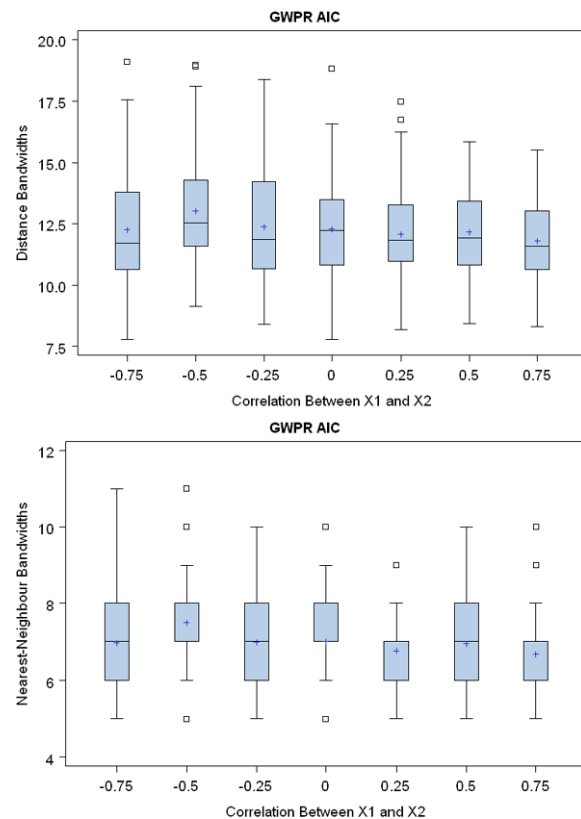
Fig. 7 Distributions of the optimal bandwidth of GWPR

Fig. 8 Distributions of the optimal bandwidth of GWNBRg

Table 3 Median of the results obtained from the GWNBR simulations, where b refers to the bandwidth and *Par* indicates the effective number of parameters

Model	Corr	b	Par	$L(\boldsymbol{\beta}, \boldsymbol{\alpha})$	AIC
PR	−0.75	–	3	−16269.11	32544.54
	−0.50	–	3	−9243.15	18492.64
	−0.25	–	3	−14632.07	29270.47
	0	–	3	−12810.82	25627.96
	0.25	–	3	−13005.57	26017.47
	0.50	–	3	−15602.13	31210.59
	0.75	–	3	−17393.98	34794.28
GWPR	−0.75	7	56.13	−391.16	1244.44
	−0.50	7	55.18	−354.44	1124.20
	−0.25	7	56.24	−420.55	1378.83
	0	7	55.85	−439.40	1359.65
	0.25	7	56.34	−404.53	1397.97
	0.50	7	56.08	−490.22	1479.71
	0.75	7	56.66	−502.84	1760.46
NBR	−0.75	–	4	−353.29	715.14
	−0.50	–	4	−349.33	707.21
	−0.25	–	4	−373.17	754.90
	0	–	4	−378.77	766.10
	0.25	–	4	−390.60	789.75
	0.50	–	4	−411.34	831.23
	0.75	–	4	−422.81	854.17
GWNBRg	−0.75	49	8.76	−1959.04	3936.19
	−0.50	52	8.25	−2003.00	4021.74
	−0.25	51	8.70	−2047.44	4110.23
	0	59	7.30	−2116.69	4259.69
	0.25	50	8.57	−1905.63	3829.91
	0.50	53	7.47	−1077.48	2167.11
	0.75	49	8.32	−1586.76	3190.46
GWNBR	−0.75	45	–	−321.51	–
	−0.50	38	–	−328.57	–
	−0.25	32	–	−362.61	–
	0	31	–	−366.78	–
	0.25	34	–	−396.87	–
	0.50	36	–	−397.72	–
	0.75	36	–	−412.07	–

gressions, see (30), and which was found by minimizing the AIC for the GWPR and GWNBRg models and minimizing the CV for the GWNBR model. The column labeled *Par* indicates the effective number of parameters, see (23) and the column labeled $L(\boldsymbol{\beta}, \boldsymbol{\alpha})$ indicates the log-likelihood.

The negative binomial regression shows a goodness of fit that is better than that obtained from the Poisson regression. This finding was expected because the simulated data were simulated according to a negative binomial distribution. In addition, allowing α to vary spatially, the best estimated model was GWNBR compared with GWNBRg, which uses a fixed α. As expected, the Poisson regression was the worst model.

Figures 9, 10 and 11 show the differences between the real and the values estimated by the GWNBR, GWPR and GWNBRg models, respectively. In summary, GWNBR was the best model, which successfully adjusted the simulated data based on a spatial negative binomial distribution. In addition, GWNBRg, despite of its biased estimate for the overdispersion parameter α also gave good results. Unfortunately, the variability was very large as was observed by Paez et al. (2011) with a small sample size. It is interesting to study this behavior for larger samples.

Figure 12 shows the sample distributions of RMSE of the parameter estimates using 700 replications for each. In general, the approach that found the best bandwidth and the type of bandwidth produced the same distribution of the parameters estimates. As shown, the distribution of RMSE for GWNBR model presents smaller values and less variability than GWNBRg and GWPR models.
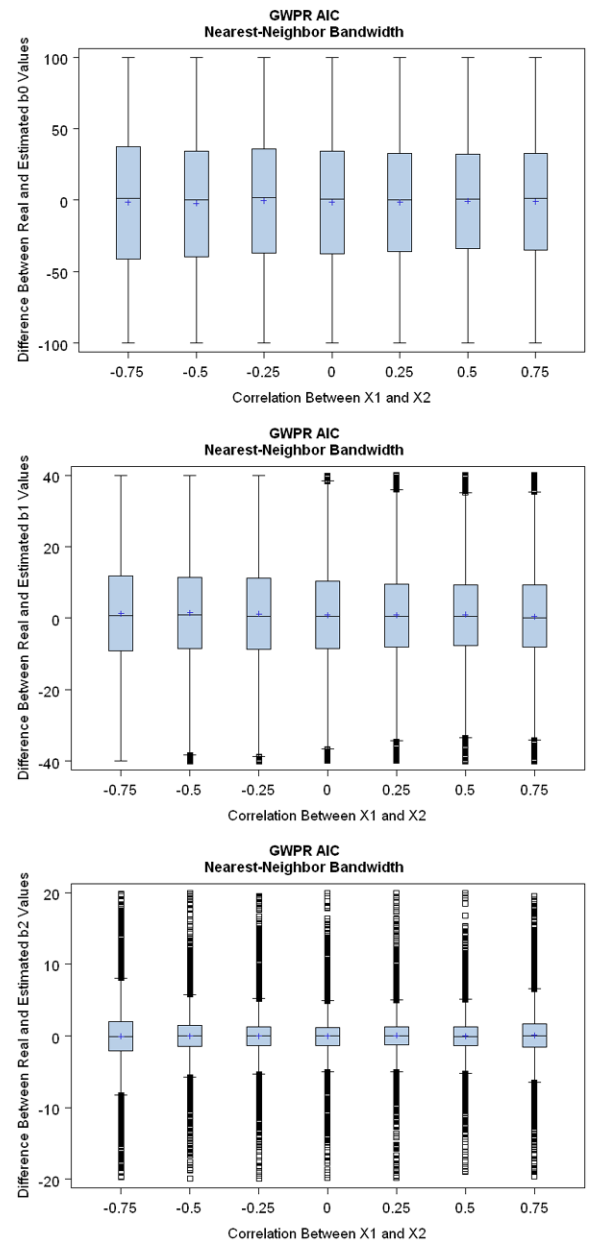
3.2 Simulation of GWPR

The main objective of this section is to show that GWNBR can also model Poisson data that are spatially distributed
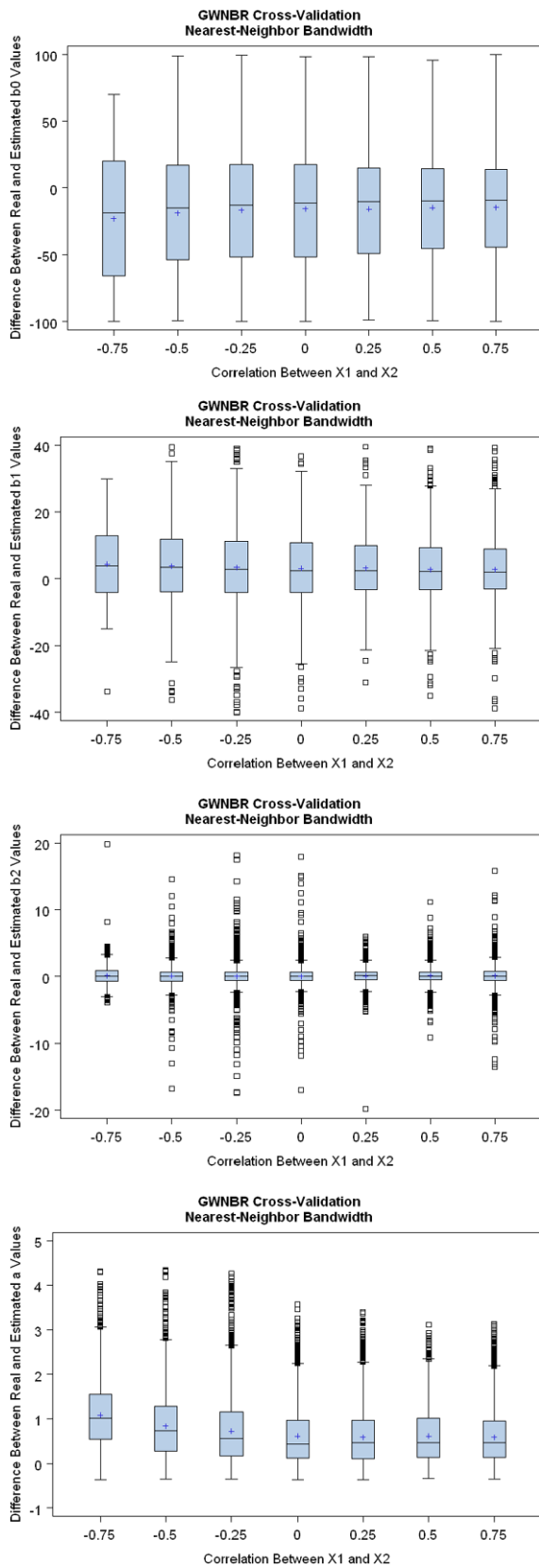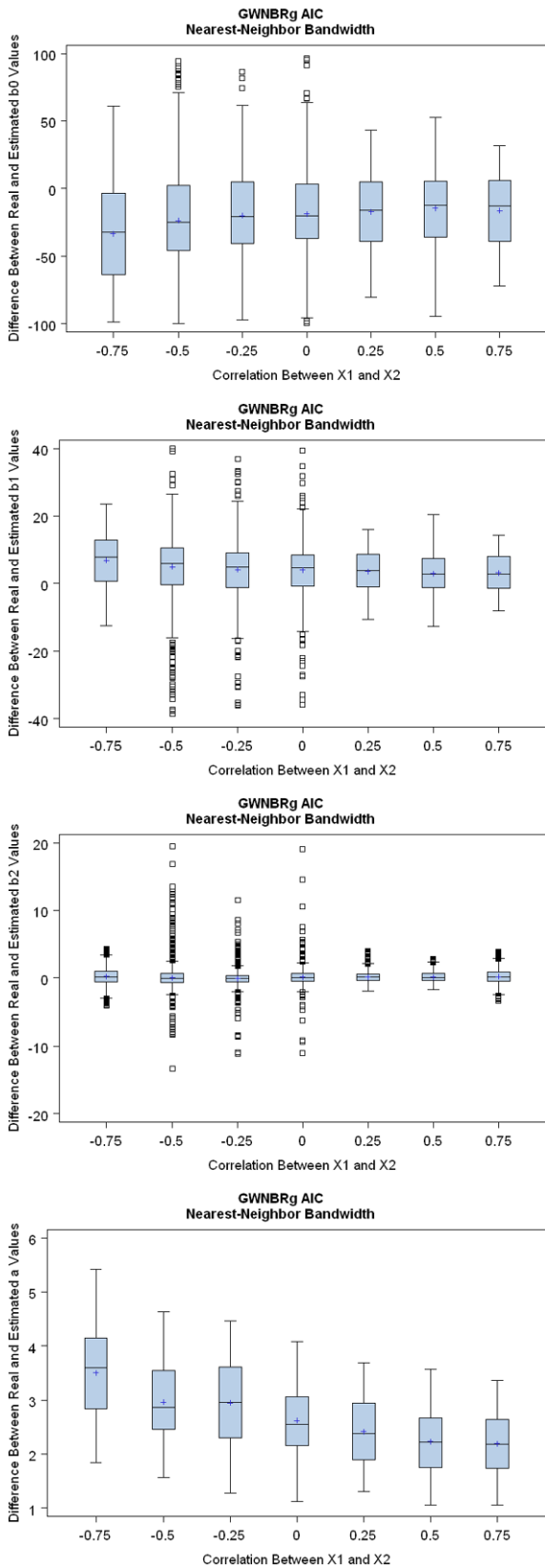
Fig. 9 Distributions of the differences between the real values and the values estimated with the GWNBR model


Fig. 10 Distributions of the differences between the real values and the values estimated with the GWPR model

and to therefore reinforce that GWPR is a special case of GWNBR. The simulated data were created 100 times, as in the previous section, assuming that α is equal to zero and that variable x_{j1} follows a normal $(0,1)$ distribution. We choose to not present all detailed results of the GWPR simulated data because of length constraints and because the Poisson distribution was very stable to the correlation between variables. The results correspond to the median values.

$$y_j \sim \text{Poisson}\big[\exp\big\{b_0(u_j, v_j) + b_1(u_j, v_j)x_{j1}\big\}\big] \qquad (32)$$

Fig. 11 Distributions of the differences between the real values and the values estimated with the GWNBRg model
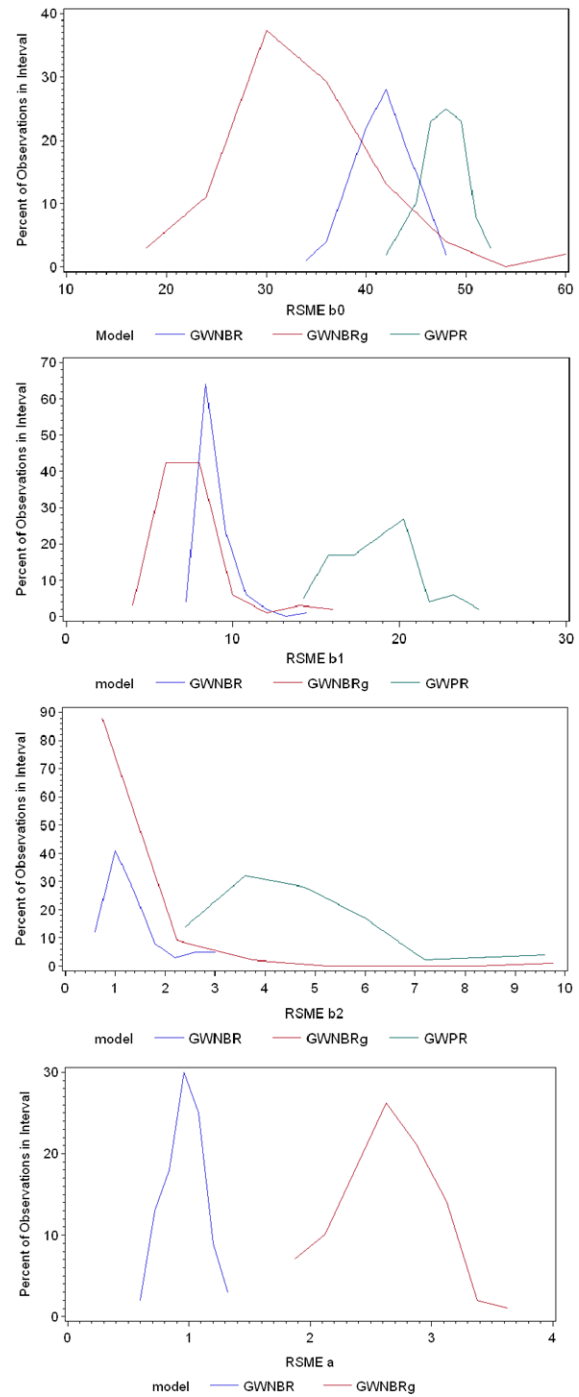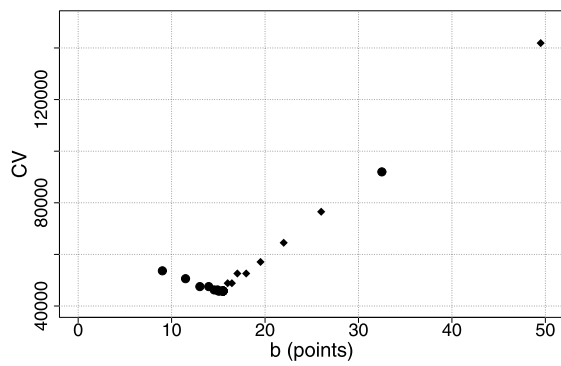

Fig. 12 Distributions of the RMSE of the parameters estimates

where

$$b_0(u_j, v_j) = 0.000005[|u_j|]^3 + 0.000005[|v_j|]^3$$
$$b_1(u_j, v_j) = \sqrt{[u_j - \overline{u}_j]^2 \times [v_j - \overline{v}_j]^2} \tag{33}$$
$$\overline{u}_j = -40.7693, \qquad \overline{v}_j = -19.5915$$

Fig. 13 Bandwidth of the GWNBR model that minimize the CV

**Table 4** Results obtained from simulations using the GWPR model

Model	b	Par	$L(\boldsymbol{\beta}, \boldsymbol{\alpha})$	AIC
PR	–	2	-1515.6	3035.3
GWPR	15	25.5	-272.8	623.5
NBR	–	3	-401.7	809.7
GWNBRg	48	7.8	-368.3	754.1
GWNBR	15	–	-282.4	–

The bandwidth was found through an adaptive bi-square kernel function, see (30) and the graphic of the minimization of the CV was achieved through a golden section search algorithm (resulting in $b = 15$), as shown in Fig. 13.

Table 4 shows some statistics for the goodness of fit obtained for the models. As expected, the GWPR shows a goodness of fit that is superior to that obtained with the other models. However, the log-likelihood of GWNBR is close to that of GWPR, which shows that this model is also a candidate. The quality measures for GWNBRg were worse than those obtained with the other models. In addition, it is interesting to see that, despite the fact that the data were derived from a Poisson distribution, the Poisson regression was the worst model because the data are spatially dependent. Even the negative binomial regression showed a better fit than the PR model, because its interpretation of the spatiality was captured in the overdispersion parameter.

Figure 14 shows the estimated surfaces of parameters b_0 and b_1 and the real maps based on (33). As shown, all of the models estimated the intercept correctly, but the adjustment of parameter b_1 by GWNBRg was the worst. In contrast, the values for parameter b_1 estimated by the GWPR and GWNBR models were almost the same. This result can be explained by the fact that the parameters $\boldsymbol{\alpha}$ of GWNBR were close to zero (varying between 0 and 0.076) and, as was found previously, the negative binomial distribution approximates a Poisson distribution when α tends to zero.

The GWNBRg model does not to differentiate between non-stationarity and overdispersion well. Because α is estimated under the assumption that all of the parameters
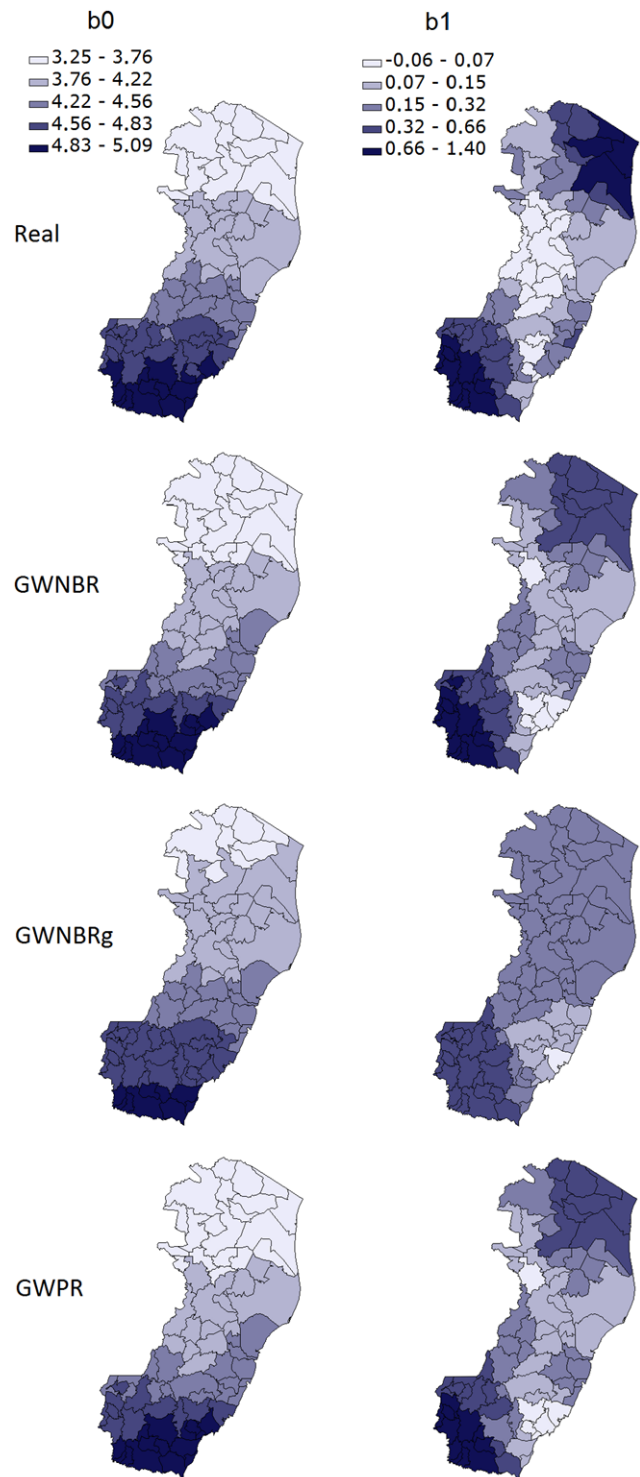
Fig. 14 Real and estimated surfaces of the parameters of the GWPR

are constants, there is confusion between the two concepts. Thus, we do not recommend that GWNBRg should be used to model count data without overdispersion.

To demonstrate that GWPR is a special case of GWNBR, we set the parameter α of the GWNBR framework to zero,

and the results, including the quality measures, such as the log-likelihood and AIC, obtained were exactly the same up to the sixth decimal point. Thus, we concluded that GW-NBR is able to model non-stationary data from a negative binomial distribution and from Poisson distribution. In addition, if $\alpha = 0$, the GWNBR framework overrides GWPR.

3.3 Application for Tokyo, Japan

The GWPR technique was developed by Nakaya et al. (2005) and used to model the working-age deaths in the Tokyo metropolitan area of Japan. The results indicate that there are significant spatial variations in the variables and that, consequently, the application of a traditional 'global' Poisson model would yield misleading results. Thus, our objective is to compare the results found by Nakaya et al. (2005) with those estimated by the GWNBR model (setting $\alpha = 0$, i.e., GWPR from GWNBR) and to also estimate the overdispersion parameter α using GWNBR.

Another objective is to determine whether GWPR is the best approach to model the mortality data from Japan and whether GWNBR is capable to produce the same results. In addition, we aimed to show the consistency of the SAS %gwnbr macro compared with the GWR 3.x software.

The explicative variables are the following:

b_1 = proportion of professional and technical workers (PRO);
b_2 = proportion of elderly people, i.e., over 64 years of age (OLD);
b_3 = rate of house-ownership (OWNH); and
b_4 = unemployment rate (UNEMP).

Table 5 shows the parameter values estimated by Nakaya et al. (2005) using the software GWR 3.x software and the parameters estimated by GWNBR using the %gwnbr macro developed in SAS/IML. Table 6 shows the standard errors of the parameters estimates. The bandwidth estimated by Nakaya et al. (2005) using the adaptive bi-square kernel given by (30) and minimizing the AIC criterion was $b = 95$ neighbors. The bandwidth estimated using GWNBR was $b = 18$ km, this bandwidth was found using the Gaussian kernel function given by (31) and minimizing the CV.

The comparison of the results shown in Table 5 and Table 6 reveals the same estimates, up to the third decimal point, between GWPR (GWR 3.x) and GWNBR (%gwnbr macro and $\alpha = 0$), which indicates that the algorithm written for GWNBR is correct.

Table 5 shows no estimates for the overdispersion parameters α for GWNBR, but these values vary between 0.0001 and 0.01 and are not significant in only 5.7 % of the areas. In addition, the log-likelihood of the GWPR and GW-NBR frameworks were close (-985.9 and -980.7, respectively), which indicates that both models exhibited a similar

Table 5 Summary of the parameter estimates obtained for the data from Tokyo, Japan

Estimates	GWR 3.x GWPR	%gwnbr GWPR	%gwnbr GWNBR
b_0			
Min	−0.942	−0.942	−0.651
Q1	−0.003	−0.003	−0.020
Median	0.099	0.099	0.095
Q3	0.259	0.260	0.161
Max	0.434	0.434	0.323
b_1			
Min	−3.783	−3.783	−4.268
Q1	−2.682	−2.681	−2.487
Median	−2.509	−2.506	−2.380
Q3	−1.800	−1.791	−1.833
Max	1.627	1.627	1.074
b_2			
Min	1.223	1.223	1.209
Q1	1.643	1.643	1.869
Median	2.069	2.079	2.066
Q3	2.437	2.448	2.401
Max	4.418	4.418	3.881
b_3			
Min	−0.572	−0.572	−0.675
Q1	−0.381	−0.380	−0.347
Median	−0.317	−0.316	−0.297
Q3	−0.210	−0.205	−0.267
Max	0.152	0.152	0.085
b_4			
Min	−0.056	−0.056	−0.057
Q1	0.021	0.021	0.028
Median	0.038	0.039	0.051
Q3	0.078	0.079	0.079
Max	0.171	0.171	0.256

goodness of fit. Thus, we conclude that the data does not exhibit overdispersion (and this fact was verified by GW-NBR). Consequently, the GWPR model is the most adequate model. Thus, to use the GWPR model with the GWNBR framework, should be set α to zero in the SAS %gwnbr macro.

3.4 Application for the Brazilian state of Espirito Santo

In this section, we analyze the distribution of the vehicles used for road freight transportation in Espirito Santo, Brazil. The understanding of the spatial distribution of these vehicles in the country can help the Brazilian authorities formulate public policies for the sector, such as supervision,
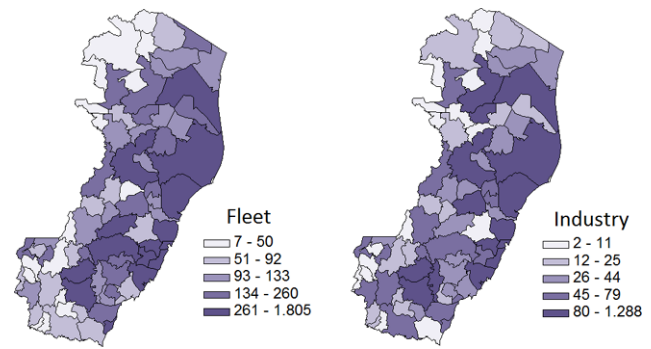
Table 6 Summary of the standard errors of the parameters estimates

Estimates	GWR 3.x GWPR	%gwnbr GWPR	%gwnbr GWNBR
Std b_0			
Min	0.094	0.094	0.106
Q1	0.128	0.128	0.120
Median	0.172	0.172	0.156
Q3	0.211	0.211	0.226
Max	0.285	0.284	0.474
Std b_1			
Min	0.227	0.225	0.233
Q1	0.306	0.305	0.277
Median	0.449	0.449	0.393
Q3	0.586	0.590	0.572
Max	0.923	0.914	1.907
Std b_2			
Min	0.310	0.309	0.296
Q1	0.375	0.375	0.342
Median	0.474	0.474	0.433
Q3	0.663	0.667	0.617
Max	0.939	0.944	1.358
Std b_3			
Min	0.079	0.079	0.078
Q1	0.102	0.102	0.091
Median	0.121	0.121	0.114
Q3	0.146	0.146	0.145
Max	0.228	0.228	0.311
Std b_4			
Min	0.014	0.014	0.017
Q1	0.019	0.019	0.019
Median	0.029	0.029	0.026
Q3	0.039	0.039	0.039
Max	0.059	0.059	0.079

Fig. 15 Spatial distribution of the variables **Fleet** and **Industry**

for the variable industry, which indicates that there is a relationship between these variables.

To check for linearity, which is required for the GWR framework, we use the linear LOESS (Cleveland and Devlin 1988) and GAM models (Hastie and Tibshirani 1990; Wood 2006). As shown in Fig. 16, the adjustment achieved with the GAM model is more sophisticated than those obtained with the LOESS and linear models. However, linearity was also shown due to the large number of data points around the line.
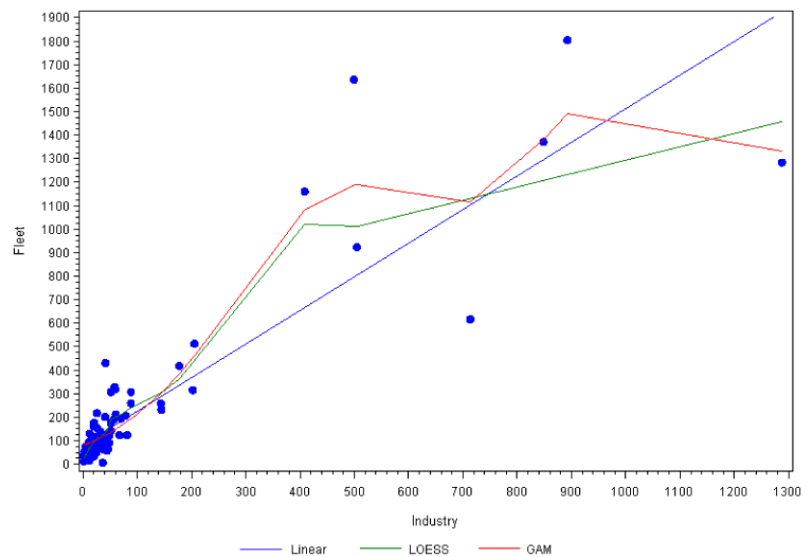
The same procedure that was followed in the analysis of the data obtained from Japan will be performed in this analysis, i.e., we will use the Poisson and Negative Binomial models, in their global and spatial forms. For the GWNBR and GWNBRg models, we used the Gaussian kernel, which resulted in a bandwidth of 53 kilometers. For the GWPR model, the bandwidth, which was found using the same kernel, was $b = 9.4$ kilometers. However, the parameter estimation with this bandwidth is inappropriate because there are only between 1 and 9 points for each regression. Therefore, we used the adaptive bi-square kernel with a bandwidth equal to 10 points. Table 7 shows the results of the fit.

As shown in Table 7, the PR and GWPR models were the worst models. The earlier problem detected in the estimation of the bandwidth was a clue of the lack of fit obtained with the Poisson distribution. Thus, we can conclude that the fleet of vehicles used in road freight transportation probably exhibits overdispersion. Therefore the negative binomial distribution is the best approach for the modeling of these variables.

Although the difference between their AIC and log-likelihood is marginal, GWNBR and GWNBRg are better choices for the modeling of these data than the NBR model. To highlight this fact, we performed a non-stationarity test based on $m = 999$ replications for the GWNBR model, by checking the variability of its estimates over space. For parameter k, the variability can be calculated as

$$V_k = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{\beta}_{ik} - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_{ik} \right)^2 \tag{34}$$

maintenance and expansion of the highways, which are the main mode of transportation in Brazil. In addition, this spatial configuration is of interest to the companies that need these vehicles to transport their products.

To facilitate the interpretation and focus on the GWNBR model, we used only one explicative variable. The amount of industries (X) in the city, which represents the economic aspect of the problem, will help explain the amount of vehicles used in road freight transportation (Y). The spatial distributions of the variables (Fleet—Y and Industry—X) are shown in Fig. 15.

As shown in Fig. 15, the trucks are concentrated in the sea border of the state (southeast side), where the capital of the state Vitoria is located. The same behavior was observed

Fig. 16 Relationship between
the variables **Fleet** and **Industry**

Table 7 Model results for the analysis of the data from Espirito Santo, Brazil

Model	b	Par	$L(\boldsymbol{\beta}, \boldsymbol{\alpha})$	AICc
PR	–	2	−5001.4	10006.9
GWPR	10 points	32.4	−1296.8	2708.2
NBR	–	3	−458.5	923.3
GWNBRg	53.07 km	8.7	−444.4	908.7
GWNBR	53.20 km	–	−440.1	–

The significance of the V_k can be evaluated through a randomization test (Hope 1968), in which the empirical distribution of the V_k is created under the null hypothesis of spatial stationarity. Initially, the geographic coordinates of the observations are randomly permuted and GWNBR is computed for this new configuration. According to the new estimates, we calculate V_k using (34), and repeat this procedure m times until the empirical distribution of V_k, under the null hypothesis is obtained. By ranking the values of V_k in descending order, including the original value for V_k we obtain the p-value using the following equation: $p/(m+1)$, where p is the rank of the original V_k, i.e., the number of times that the simulated estimates were greater than the original estimate. The randomization test has the advantage that it does not use a probability distribution for V_k, but it is computationally intensive. Another parametric test for non-stationarity was proposed by Leung et al. (2000a), but this method was not used in this analysis.

The p-values for β_0, β_1, and α were 1.2 %, 8.4 % and 62.6 %, respectively. Considering the null hypothesis of stationarity, we can conclude that β_0 and β_1 are statistical different across the space at level of 10 %.

The surface of the parameters and their standard errors are shown in Fig. 17. The greater values for the intercept (b_0)

are concentrated in the Vitoria metropolitan region (southeast side), which is the capital of the state. This finding reflects the greater amount of vehicles located in that place. In contrast, the parameter estimates for b_1 are smaller in that area because there are many industries and the link function is exponential.

In summary, the results indicate that there is spatial variation and overdispersion in the distribution of vehicles used for road freight transportation in Espirito Santo, Brazil. Consequently, the application of a traditional negative binomial model would yield misleading results. In addition, the Poisson models (global or spatial) would produce even poorer results.

4 Conclusions

This paper aims to show a methodology for Geographical Weighted Regression for data that follow a negative binomial distribution; this method is denoted GWNBR. Due to the difficulty associated with finding the effective number of parameters, we also developed the GWNBRg approach. The difference between these two methods is in the estimation of the overdispersion parameter α. In GWNBRg, α is estimated in globally, which generates a biased estimate for the parameter of overdispersion. However, the simplicity of this approach allows the calculation of the effective number of parameters of the model. In GWNBR, although this amount is unknown, GWNBR allows parameter α to vary spatially.

Based on the simulations presented, it was found that the global regressions search for regularities and patterns is inadequate for the search for local particularities or exceptions. In addition, the models based on the Poisson distribution were found to be inappropriate for the modeling of data with overdispersion. Also, it is interesting to investigate
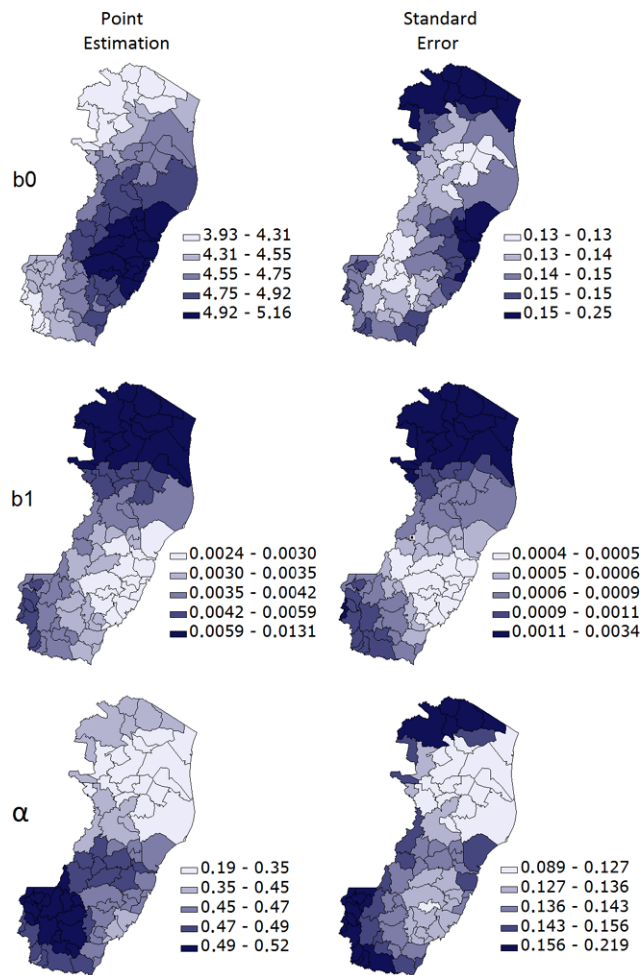
Fig. 17 Surface of the parameter estimates and standard errors obtained with the GWNBR model

whether the distribution of α influenced the results, i.e, simulate a spatial distribution of α for data that do not exhibit overdispersion or non smoothy (more rugged data).

GWNBR shows a goodness of fit in the modeling of non-stationary and overdispersion data that is superior to that obtained with the other models, as was observed in the example of the fleet of vehicles in the Brazilian state of Espírito Santo. Also, GWNBR exhibits a good approximation for GWPR when $\alpha \to 0$, or the same results when $\alpha = 0$, as was observed in the example of the distribution of deaths in Tokyo, Japan.

The simulations of count data showed a pattern that was the same as that found by Paez et al. (2011), who evaluated the results of Gaussian GWR in the presence of multicollinearity, i.e., when the data are created from a negative binomial distribution, NBR but not PR is sensitive to multicollinearity. The simulations also showed that the approach used to find the best bandwidth influences the results: AIC has a smaller variability than CV but CV produces a better fit than AIC, as measured by the log-likelihood.

In conclusion, the GWNBR method is a robust tool for the modeling of count data, especially when the data exhibit non-stationarity and overdispersion, by incorporating the Poisson Regression (PR), Geographically Weighted Poisson Regression (GWPR), and Negative Binomial Regression (NBR) models.

References

Casella, G., Berger, R.L.: Statistical Inference, 2nd edn. Duxbury, N. Scituate (2001)

Cleveland, W.S., Devlin, S.J.: Locally-weighted regression: and approach to regression analysis by local fitting. J. Am. Stat. Assoc. **83**, 596–610 (1988)

Farber, S., Paez, A.: A systematic investigation of cross-validation in gwr model estimation: empirical analysis and Monte Carlo simulations. J. Geogr. Syst. **9**(4), 371–396 (2007)

Farber, S., Paez, A., Volz, E.: Topology and dependency tests in spatial and network autoregressive models. Geogr. Anal. **41**(2), 158–180 (2009)

Fotheringham, A.S., Brunsdon, C., Charlton, M.: Geographically Weighted Regression. Wiley, New York (2002)

Geniaux, G., Ay, J.S., Napoléone, C.: A spatial hedonic approach on land use change anticipations. J. Reg. Sci. **51**(5), 967–986 (2011)

Hadayeghi, A., Shalaby, A.S., Persaud, B.N.: Development of planning level transportation safety tools using geographically weighted Poisson regression. Accid. Anal. Prev. **42**, 678–688 (2010)

Hastie, T.J., Tibshirani, R.J.: Generalized Additive Models. Chapman & Hall/CRC, London (1990)

Hilbe, J.M.: Negative Binomial Regression, 2nd edn. Cambridge University Press, Cambridge (2011)

Hope, A.C.A.: A simplified Monte Carlo significance test procedure. J. R. Stat. Soc. **30**, 582–598 (1968)

Kobayashi, T., Lane, B.: Spatial heterogeneity and transit use. In: 11th World Conference on Transportation Research (2007)

Leung, Y., Mei, C.L., Zhang, W.X.: Statistical tests for spacial nonstationarity based on the geographically weighted regression model. Environ. Plan. A **32**, 9 (2000a)

Leung, Y., Mei, C.L., Zhang, W.X.: Testing for spatial autocorrelation among the residuals of the geographically weighted regression. Environ. Plan. A **32**, 871–890 (2000b)

Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M.: Geographically weighted Poisson regression for disease association mapping. Stat. Med. **24**, 2695–2717 (2005)

Nelder, J.A., Wedderburn, R.W.M.: Generalized linear models. J. R. Stat. Soc. **135**, 370–384 (1972)

Paez, A., Farber, S., Wheeler, D.: A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. Environ. Plan. A **43**(12), 2992–3010 (2011)

Staniswalis, J.G.: The kernel estimate of a regression function in likelihood-based models. J. Am. Stat. Assoc. **84**, 276–283 (1989)

Wheeler, D., Tiefelsdorf, M.: Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J. Geogr. Syst. **7**(2), 161–187 (2005)

Wood, S.: Generalized Additive Models: An Introduction with R, 2nd edn. Chapman & Hall/CRC, London (2006)