

Regressão linear geograficamente ponderada em ambiente SIG

Luís Eduardo Ximenes Carvalho¹; Henrique Nogueira Silva²;
Carlos Felipe Grangeiro Loureiro³; Hamifrancy Brito Meneses⁴

Resumo: Este artigo aborda considerações teóricas e resultados da implementação em ambiente SIG de um modelo confirmatório de estatística espacial — regressão linear geograficamente ponderada (RGP) — não disponível em ambiente livre. Os aspectos teóricos deste modelo local de regressão espacial foram amplamente discutidos em virtude da escassa bibliografia existente. O modelo RGP foi implementado na linguagem de programação GISDK do SIG-T TransCAD, utilizando compreensivamente as ferramentas de manipulação, tratamento georreferenciado dos dados e rotinas de análise espacial disponibilizadas em plataformas SIG. Ao final, espera-se ter desenvolvido, ainda que de maneira parcial, uma importante ferramenta que contribuirá para a compreensão e refinamento da modelagem de fenômenos geográficos tão amplamente analisados em estudos de Planejamento de Transportes.

Abstract: This paper discusses theoretical issues and results of the implementation in a GIS environment of a spatial statistic confirmatory model — geographically weighted linear regression (GWR) — not available in a free software environment. The theoretical aspects of this local spatial regression model were vastly discussed due to the lack of related bibliography. The GWR model was coded in TransCAD's programming language, named GISDK, making a comprehensive use of manipulation tools, data geo referenced treatment, and spatial analysis routines already existing in GIS platforms. At the end, we hope to have developed, although in a partial manner, an important tool which will contribute to a better understanding and fine tuning of geographical phenomena modeling so widely applied in transportation planning studies.

1. INTRODUÇÃO

Em razão da larga utilização de modelos de regressão linear para a descrição e predição de inúmeros fenômenos físicos, este relativo sucesso tem motivado a aplicação desta técnica em outras áreas do conhecimento, especialmente nas ciências econômicas. Para o caso de estudos de Planejamento de Transportes, Lopes (2005) cita o amplo emprego da regressão linear convencional na etapa de geração de viagens da modelagem da demanda. Porém, a simples aplicação deste modelo estatístico em fenômenos com caráter eminentemente geográfico, como é o caso daqueles relacionados aos sistemas de transportes, pode induzir a presença de associações espúrias e superestimação da significância dos seus parâmetros (Câmara *et al.*, 2002).

Buscando representar melhor o comportamento da variável em análise a partir das distribuições de outras variáveis associadas, diversos modelos confirmatórios capazes de incorporar o atributo espacial — comumente denominados de *modelos de regressão espacial*

— têm sido desenvolvidos sob a expectativa de refinamento na descrição e predição dos fenômenos geográficos (Fotheringham *et al.*, 2000). Entretanto, na escolha do tipo de modelo espacial mais adequado, deve-se considerar que a elevação no poder de explicação proporcionada pela regressão espacial sobre a convencional está condicionada ao grau de dependência espacial do fenômeno analisado (Silva, 2006). Além disso, a verificação do tipo e intensidade da autocorrelação da variável analisada no espaço geográfico, assim como da sua correlação com as variáveis explicativas, está diretamente relacionada à presença ou não de estacionariedade espacial, isto é, relações de dependência constantes sobre o espaço. Se o fenômeno em análise apresentar forte estacionariedade, modelos globais de regressão espacial, com coeficientes únicos para toda a área em estudo, serão capazes de descrever bem a relação observada de dependência espacial. Caso contrário, essa representação global pode levar a estimativas pouco precisas em alguns pontos, justificando-se a calibração de coeficientes específicos para diferentes regiões do espaço, o que significa estimar modelos locais de regressão espacial. Nesta categoria, destaca-se a *regressão linear geograficamente ponderada*, ou de forma abreviada, *regressão linear geponderada*, foco deste trabalho.

Embora a aplicação de modelos de regressão espacial com efeitos globais e locais tenha um vasto potencial no tratamento da problemática dos transportes urbanos e regionais, uma parcela desses modelos está concebida apenas conceitualmente, enquanto outros ainda não estão disponíveis em pacotes computacio-

¹ Luís Eduardo Ximenes Carvalho, Brown University, Providence, Rhode Island, EUA. (e-mail: carvalho@dam.brown.edu).

² Henrique Nogueira Silva, Universidade Federal do Ceará. Fortaleza, CE, Brasil. (e-mail: henriquensilva@gmail.com).

³ Carlos Felipe Grangeiro Loureiro, Universidade Federal do Ceará. Fortaleza, CE, Brasil. (e-mail: felipe@det.ufc.br).

⁴ Hamifrancy Brito Meneses, Centro Federal de Educação Tecnológica do Ceará. Fortaleza, CE, Brasil. (e-mail: brito@det.uf.br).

nais em ambiente livre. Ademais, não se tem conhecimento de qualquer desses modelos implementados em plataformas computacionais de Sistemas de Informações Geográficas (SIG). Portanto, este artigo tem como propósito abordar os aspectos teóricos e conceituais que embasam a formulação do modelo de regressão linear geograficamente ponderada, assim como descrever sua implementação na linguagem de programação GISDK do SIG-T TransCAD, utilizando compreensivamente as ferramentas de manipulação, tratamento georreferenciado dos dados e rotinas de análise espacial disponibilizadas em plataformas SIG. Em seqüência a este trabalho, Loureiro *et al.* (2006) propõem uma metodologia de calibração deste tipo de modelo, tendo como exemplo de aplicação a análise da geração de viagens radiais (com destino a Fortaleza) no Sistema de Transporte Rodoviário Intermunicipal do Estado do Ceará.

2. ASPECTOS TEÓRICOS DA REGRESSÃO GEOGRAFICAMENTE PONDERADA

A *regressão linear geograficamente ponderada (RGP)* foi inicialmente proposta por Brunson *et al.* (1996) como um método para explorar a não-estacionariedade espacial, sendo esta última uma condição na qual um modelo de regressão espacial global não pode, conforme destacado acima, explicar adequadamente as relações entre alguns conjuntos de variáveis definidas numa região geográfica. Deste modo, uma solução mais adequada seria modificar a natureza do modelo para refletir, ao longo do espaço, a estrutura intrínseca aos dados.

A idéia básica da RGP é ajustar um modelo de regressão para cada ponto no conjunto de dados, ponderando as observações por uma função de distância a este ponto. Isto corresponde a considerar que pontos mais próximos ao ponto em estudo tenham maior influência nos parâmetros estimados da regressão do que observações obtidas em pontos mais distantes. Como resultado, obtêm-se portanto um conjunto de parâmetros ajustados para cada ponto na região geográfica analisada. A proposta inicial tinha como objetivo fornecer uma ferramenta útil de análise exploratória através da identificação das variações entre variáveis ao longo do espaço, o que corresponde à não-estacionariedade espacial; entretanto, por auxiliar na explicação e previsão de fenômenos geográficos, este tipo de modelo pode ser classificado como confirmatório, desde que, conforme discutido no trabalho de Loureiro *et al.* (2006), seja avaliada com o devido cuidado sua premissa básica de influência do comportamento de uma determinada unidade geográfica sobre as suas vizinhas, e vice-versa. A seguir, apresenta-se uma descrição mais detalhada deste modelo, bem co-

mo considerações sobre os diversos tipos de ponderações geográficas passíveis de utilização na RGP.

2.1. Concepção do modelo RGP

Considerando inicialmente um modelo convencional de regressão linear multivariável não-espacial, no qual uma variável dependente y é representada como uma combinação linear de variáveis explicativas x_k , pode-se obter uma estimativa de y para um ponto i a partir da equação de regressão:

$$y_i = a_0 + \sum_{k=1}^p a_k \cdot x_{k,i} + \varepsilon_i \quad (1)$$

em que,

- $x_{k,i}$: k -ésima variável explicativa na localização i ;
- a_0 : intercepto global (valor constante);
- a_k : k -ésimo coeficiente angular global referente a x_k (valor constante);
- ε_i : erros independentes.

Para estimativa dos coeficientes angulares e do intercepto em um modelo de regressão linear convencional, normalmente utiliza-se o método dos quadrados mínimos simples, resultando na seguinte equação dos estimadores em notação matricial:

$$\hat{A} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y \quad (2)$$

em que,

- \hat{A} : vetor dos coeficientes (intercepto e coeficientes angulares);
- X : matriz de observações das variáveis explicativas;
- Y : vetor de observações da variável dependente (explicada).

Vale observar, portanto, que para um conjunto de observações é obtido o mesmo conjunto de estimativas para os coeficientes na Equação 1. Logo, se as observações forem coletadas para um conjunto de pontos, os estimadores obtidos serão constantes ao longo do espaço. A regressão geograficamente ponderada (RGP) consiste numa técnica simples de extensão da regressão convencional permitindo que parâmetros locais – em vez de globais – sejam estimados (Fotheringham *et al.*, 1997). Deste modo, as estimativas tornam-se específicas para cada localização i , e o modelo pode então ser reescrito como:

$$y_i = a_{0,i} + \sum_{k=1}^p a_{k,i} \cdot x_{k,i} + \varepsilon_i \quad (3)$$

em que,

- $a_{0,i}$: intercepto local específico da localização i ;
- $a_{k,i}$: valor do k -ésimo coeficiente angular local específico da localização i .

Pode-se notar, a partir das Equações 1 e 3, que a principal diferença entre os modelos está na extensão e generalidade obtida para cada localização i , passando de observação para um modelo específico, particu-

lar. Assumindo que os parâmetros exibam algum grau de consistência espacial, pode-se considerar, por exemplo, um subconjunto de localizações próximas a que está sendo estudada como fornecedor de informações adicionais. Como este tipo de distribuição espacial de valores é o mais comumente encontrado, com valores pertencentes a localizações próximas tendo grandezas e sinais semelhantes, então este será usado ao longo da discussão.

Uma abordagem geral, ao se tratar este tipo de modelo, é notar que, apesar da obtenção de estimativas não tendenciosas ser impossível — já que os coeficientes mostrarão um desvio ao longo do subconjunto local de calibração — pode-se ainda conseguir uma tendenciosidade pequena. De fato, como Fotheringham *et al.* (1997) comentam, o processo de modelagem em RGP pode ser visto como um equilíbrio de troca entre tendenciosidade e erro padrão. Quanto maior for o subconjunto local de calibração, menores serão os erros padrões das estimativas dos coeficientes; no entanto, deve-se atentar para o fato de que quanto maior o subconjunto local, maior a probabilidade de que os desvios introduzidos pela região espacial acrescentem tendenciosidade. Uma forma de reduzir tal efeito é considerar uma *função de ponderação espacial*, de modo que localizações mais distantes tenham menor influência na calibração do modelo em estudo.

De um modo geral, no *método de quadrados mínimos ponderados*, fatores de ponderação são aplicados em cada diferença ao quadrado antes da minimização, de forma que a imprecisão de alguns preditores receba maior penalidade que outros. Os estimadores ponderados para a Equação 1 podem ser então escritos como:

$$\hat{A} = (X^T \cdot W \cdot X)^{-1} \cdot X^T \cdot W \cdot Y \quad (4)$$

em que W é uma matriz quadrada cujos elementos da diagonal principal correspondem aos pesos de cada observação e os outros elementos são nulos, ou seja:

$$W = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} \quad (5)$$

Vale notar que, no modelo de regressão não-espacial (Equação 1), os pesos atribuídos a cada observação são também constantes, e logo a estimativa dos parâmetros é a mesma para todo o espaço — estimativa global. Na RGP, os quadrados mínimos ponderados são utilizados para ponderação de cada observação de acordo com sua proximidade ao ponto i , de forma que, como no modelo representado pela

Equação 3, os estimadores variem de acordo com a localização do ponto i em estudo pela variação de W . Desta forma, o estimador para os parâmetros da Equação 3 é semelhante ao da Equação 4, levando em consideração, contudo, que a matriz de ponderação depende de i . Logo, tem-se:

$$\hat{A}_i = (X^T \cdot W_i \cdot X)^{-1} \cdot X^T \cdot W_i \cdot Y \quad (6)$$

em que, analogamente, W_i é uma matriz quadrada cujos elementos da diagonal principal correspondem aos pesos de cada observação *com relação ao ponto i* e os elementos restantes são nulos. De modo semelhante a \hat{A}_i , considerando-se agora pesos para cada ponto i , deve-se estender também a notação. Por $w_{i,j}$ denota-se o peso atribuído à observação (ponto) j em relação ao ponto i . Logo,

$$W_i = \begin{bmatrix} w_{i,1} & 0 & \cdots & 0 \\ 0 & w_{i,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{i,n} \end{bmatrix} \quad (7)$$

2.2. Funções de ponderação espacial e processo de calibração da RGP

Pode-se notar pela Equação 6 que, uma vez fornecidos os parâmetros de entrada X e Y , basta a definição das matrizes W_i para completar o modelo de regressão geoponderada. Isto pode ser feito através da escolha adequada de *funções de ponderação espacial*, como já anteriormente mencionado.

Considerando inicialmente o modelo de regressão não-espacial estabelecido pelas Equações 1 e 2, este pode ser visto como um caso particular da Equação 6, na qual:

$$w_{i,j} = 1 \quad \forall i, j \quad (8)$$

onde j representa um ponto específico no espaço no qual os dados são observados e i um ponto no qual os parâmetros são estimados. Aqui todas as observações têm peso unitário, o que corresponde a uma estimativa constante dos parâmetros ao longo do espaço, ou seja, ao modelo de regressão global já discutido anteriormente. Uma forma de considerar pesos consistentes com a proximidade e, portanto, um modelo de características locais, seria excluir do processo de calibração local observações que estejam mais distantes do ponto em estudo do que uma determinada distância de influência ($D_{influencia}$). Isto seria equivalente a fornecer valores nulos para tais observações, o que corresponde à função de ponderação:

$$w_{i,j} = \begin{cases} 1 & \text{se } d_{i,j} < D_{influencia} \\ 0 & \text{caso contrário} \end{cases}, \quad \forall i, j \quad (9)$$

em que $d_{i,j}$ representa a distância entre os pontos i e j (Fotheringham *et al.*, 1997).

Embora a consideração de $w_{i,j}$, de acordo com a Equação 9, forneça um modelo de regressão local e geograficamente ponderado, este sofre do problema de *descontinuidade*. Como Fotheringham *et al.* (1997) observam, à medida que i varia ao longo da região em estudo, os coeficientes das regressões podem mudar drasticamente, já que os pontos de observação podem estar incluídos ou não na região circular em torno de i , ou seja, o modelo é fortemente dependente do parâmetro $D_{influencia}$. Uma forma de contornar tal problema é especificar $w_{i,j}$ como uma função *contínua* da distância $d_{i,j}$. Uma escolha razoável seria considerar uma curva gaussiana:

$$w_{i,j} = e^{(-\beta \cdot d_{ij}^2)} \quad (10)$$

em que β é um parâmetro de decaimento de acordo com a distância. A vantagem da consideração de funções contínuas está na inclusão de contribuições fracionárias das observações, de modo que pontos mais distantes tenham contribuições menos significativas. Vale observar que pesos fracionários *não* alteram o valor da observação, mas sim sua influência no processo de calibração do modelo. Para observações em pontos muito distantes de i , os pesos são praticamente nulos, efetivamente excluindo tais observações. Uma forma de conciliar as funções de ponderação, representadas nas Equações 9 e 10, é estimar β com base no valor crítico $D_{influencia}$ e um valor de *tolerância* e .

$$\beta = \frac{-\ln(e)}{D_{influencia}^2} \quad (11)$$

onde e representa a proximidade a zero desejada. Um valor razoável seria $e = 0,0001$ que corresponde a atribuição de peso virtualmente igual a zero para as observações do ponto j quando este dista linearmente de $D_{influencia}$ do ponto i em análise. O problema agora da calibração do modelo RGP reside na escolha adequada de valores para β ou, alternativamente, para $D_{influencia}$. Na medida em que β tende a zero — ou, equivalentemente, $D_{influencia}$ tende a maior distância entre i e j — os pesos tendem ao valor unitário para todos os pares i,j de pontos, de forma que os parâmetros estimados tornam-se uniformes, resultando em um modelo global. Por outro lado, à medida que β torna-se maior, as estimativas dos parâmetros dependerão cada vez mais das observações mais próximas de i , e, conseqüentemente, terão maior variância.

Fotheringham *et al.* (2000) sugerem um processo de *validação cruzada*, no qual são utilizados os valores,

$$SQE = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(\beta)]^2 \quad (12)$$

em que,

- SQE : soma dos quadrados dos erros;
- y_i : valor observado da variável explicativa no ponto i ;
- $\hat{y}_{\neq i}$: valor ajustado de y_i de acordo com $(\beta$ ou $D_{influencia})$ com o ponto i omitido do processo de calibração.

A escolha final destes parâmetros (β ou $D_{influencia}$) geralmente será baseada na busca da minimização de SQE do fenômeno em estudo. É importante ainda chamar a atenção que SQE, embora utilizado no modelo local RGP, é um parâmetro global que, em função desta característica, servirá para comparar globalmente o nível de desempenho da RGP em relação aos modelos convencionais e espaciais globais. Além disso, a associação deste parâmetro com a variabilidade total da variável dependente y permite avaliar coeficientes globais de determinação (R^2) que também servirão para comparações.

Vale também destacar que, por questões de convergência do modelo RGP, o ponto i em análise é excluído da calibração de sua própria reta de regressão. A não consideração desta medida resultará na minimização de SQE para uma faixa de valores demasiadamente elevados para β (ou, equivalentemente, para uma faixa de valores de $D_{influencia}$ demasiadamente pequenos), produzindo uma função de ponderação (Equação 10) bastante acentuada, de tal forma que os pesos atribuídos aos pontos observados j 's serão virtualmente iguais a zero, não exercendo influência sobre o ponto em análise i e fazendo com que o valor predito y_i dependa *exclusivamente* da observação do ponto i , o que obviamente não é desejado.

3. IMPLEMENTAÇÃO DA RGP EM AMBIENTE SIG

A implementação da regressão geograficamente ponderada (RGP) foi realizada em ambiente SIG, codificando uma rotina computacional no pacote de SIG-T TransCAD 4.5 (Caliper, 2002a). Esta rotina foi desenvolvida no sentido de contemplar os conceitos da RGP descritos da seção anterior. Para tanto, utilizou-se a linguagem de programação GISDK (Caliper, 2002b), específica deste ambiente, cuja implementação foi vantajosamente realizada devido à visualização gráfica e à classificação temática dos *inputs* que permitiram melhor análise dos resultados. Além do mais, a implementação em si permitiu usar bancos de dados georreferenciados para tratamento das observações e operações topológicas como, por exemplo, a pesquisa proximal.

Tabela 1: Características da RGP e da Regressão Não-Espacial

Característica	Regressão Espacial - RGP -	Regressão Não-Espacial - Convencional -
Considera o atributo espacial dos dados	sim	não
Natureza do modelo	local	global
Modelos não-lineares	não implementado	disponível
Funções de ponderação geográfica	implementado (tipos constante e gaussiana)	não se aplica
Coefficientes r e R ² global	implementado	disponível
Coefficientes angulares locais	implementado	não se aplica
Desconsideração do intercepto	implementado	disponível
Coefficientes r e R ² local	não implementado	não se aplica
Análise de significância dos parâmetros	não implementado	disponível

A Tabela 1 descreve as principais informações relativas aos aspectos já implementados e ainda por implementar da RGP neste ambiente. Para fins de comparação, são mostradas também as características disponíveis da regressão não-espacial (convencional), bem como aquelas que não se aplicam em função da concepção global deste modelo.

É importante ressaltar que mesmo no caso de fenômenos intrinsecamente espaciais, conforme destacado por Loureiro *et al.* (2006), a calibração preliminar de modelos de regressão não-espacial se constitui como uma boa prática para fornecimento de informações adicionais, como, por exemplo, a análise de multicolinearidade entre variáveis explicativas x_k . Como se pode observar na Tabela 1, existem características da RGP ainda não implementadas neste ambiente de desenvolvimento. Porém, a implementação realizada já permite aplicar satisfatoriamente a RGP para obtenção dos coeficientes locais do modelo. Nos itens seguintes, apresenta-se uma descrição das características implementadas no SIG.

3.1. Modelagem geográfica da RGP

A modelagem da RGP é realizada com base em informações de unidades geográficas, que na maioria

dos estudos de planejamento de transportes são representadas em um SIG por áreas, como por exemplo, setores censitários, zonas de tráfego ou municípios. Porém, para fins de mensuração da interação espacial entre unidades geográficas, foi estabelecida uma relação unívoca entre a camada de área e uma camada de pontos. Na falta de informações mais precisas, considera-se o centróide da área como representativo; porém, deve-se sempre investigar se existe algum ponto mais representativo da unidade geográfica (caso dos pontos que representam as sedes de municípios que não necessariamente coincidem com o centróide).

Com o intuito de ilustrar a implementação realizada, desta seção em diante serão exibidos dados referentes aos estudos em desenvolvimento da mobilidade urbana de Fortaleza (Henrique *et al.*, 2005), com dados agregados no nível de zonas de tráfego com seus respectivos centróides (Figura 1). O banco de dados georreferenciado da Figura 1, além de possuir informações da variável dependente y e das variáveis explicativas x , únicos dados de entrada necessários para calibração de modelos de regressão não-espacial, considera informações de *longitude* e *latitude* que serão necessárias para avaliar a localização geográfica do pon-

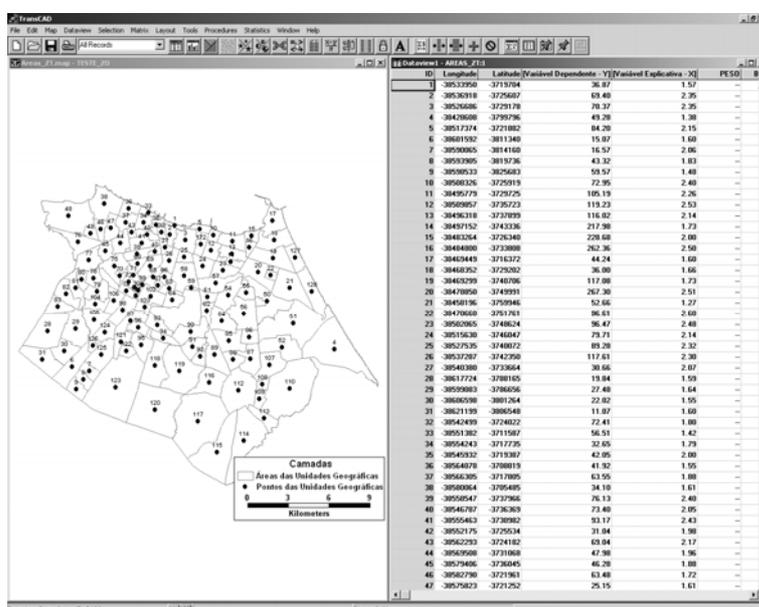


Figura 1: Dados de entrada da RGP – camada geográfica de pontos e banco de dados georreferenciado

to para processamento das informações espaciais — como a distância entre unidades geográficas, por exemplo.

Os demais parâmetros de entrada são fornecidos por meio de uma caixa de diálogo (Figura 2) em que devem ser informados o nome da camada de pontos, as distâncias de influência inicial ($D_{influencia,i}$) e final ($D_{influencia,f}$), bem como o passo de iteração (p) da pesquisa de minimização dos resíduos. Estas distâncias de influência devem ser avaliadas conforme a escala do fenômeno em estudo, recomendando-se como valor mínimo (inicial) a distância linear entre os centróides do par de unidades geográficas vizinhas mais afastadas. Para o valor máximo (final), recomenda-se a maior dimensão da área em estudo. Deve-se informar ainda, para o caso da ponderação gaussiana (ver Equações 10 e 11), a precisão escolhida “ e ” que corresponde ao peso da unidade geográfica na calibração da equação de ajuste do ponto considerado, quando estas distam linearmente de uma distância $D_{influencia}$.

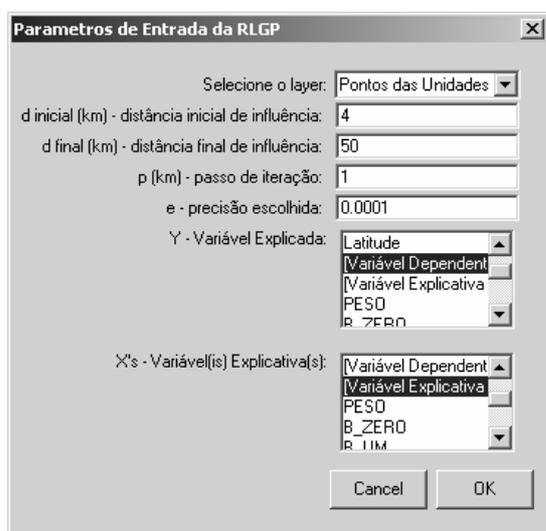


Figura 2: Caixa de diálogo dos parâmetros de entrada da RGP.

É importante chamar a atenção que, nesta implementação, se considerou apenas a distância euclidiana entre pontos, deixando como recomendação introduzir alguma distância pela rede viária ou de transporte público, ou qualquer outra medida de impedância (como tempo de viagem), traduzindo melhor a relação espacial entre unidades geográficas com considerações do nível de serviço da infra-estrutura de transportes.

3.2. Funções de ponderação espacial e desconsideração do intercepto

A rotina implementada considera dois tipos de funções de ponderação geográfica, a ser escolhido pelo usuário: do tipo constante, definida pela Equação 9, e do tipo gaussiana, expressa pelas Equações 10 e 11. Para a primeira função de ponderação citada é possível inclusive uma visualização gráfica da lógica da RGP na forma de um diagrama de dispersão. A Figura 3(a) exibe o diagrama de dispersão do exemplo utili-

zado, tomando por base todas as unidades geográficas, ou seja, sem nenhuma consideração do atributo espacial; sendo, portanto, um modelo não-espacial. Na Figura 3(b), estima-se a reta de regressão local da unidade geográfica em destaque (◇) considerando somente municípios com uma distância linear inferior a $D_{influencia} = 4$ km, que é equivalente a admitir pesos nulos para unidades geográficas distantes, conforme Equação 9.

Pode-se perceber que, quando da consideração do atributo espacial, os parâmetros (coeficiente angular e intercepto) apresentaram um ajuste local à observação em destaque. A ponderação gaussiana foi a única função de ponderação espacial fracionária implementada sob a expectativa de suavização dos coeficientes modelados. Diggle (1998) cita outras funções de ponderação fracionárias que seguem distribuições de Poisson e binomial.

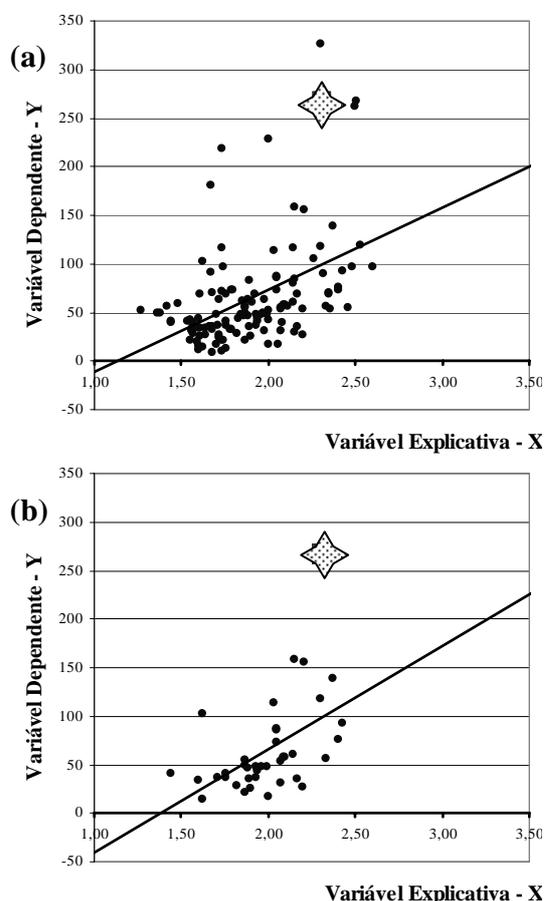


Figura 3: Diagramas de dispersão do algoritmo RGP – função de ponderação geográfica constante

De forma análoga ao modelo de regressão não-espacial, o usuário pode optar pela desconsideração do intercepto das retas locais de regressão, pois algumas vezes o valor nulo de intercepto é mais adequado para representar conceitualmente a relação de dependência entre as variáveis.

3.3. Processamento da rotina RGP

Após a escolha do tipo de ponderação (gaussiana ou constante) e definição da existência de intercepto, o

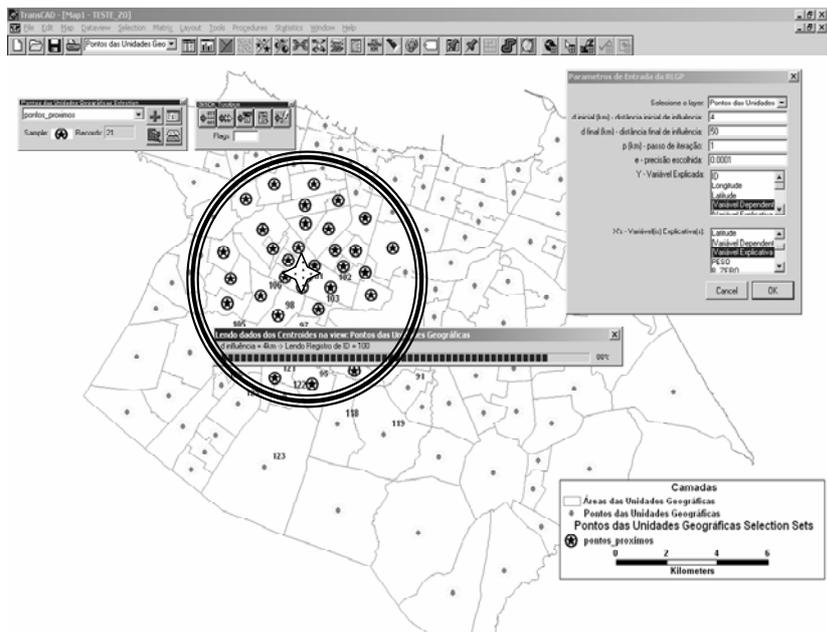


Figura 4: Execução da rotina RGP no SIG TransCAD

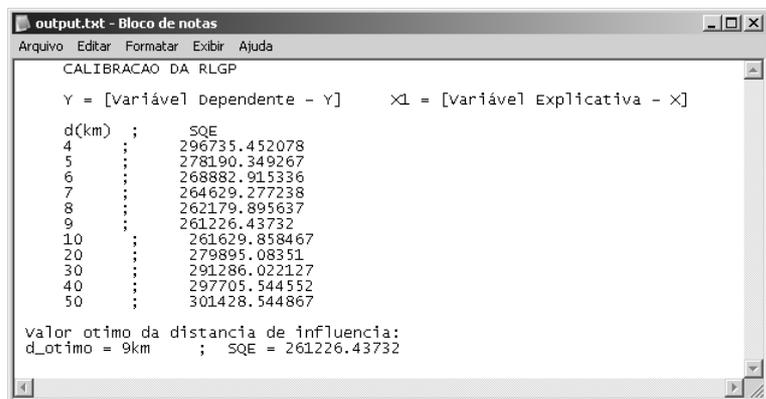


Figura 5: Arquivo de saída da RGP para pesquisa de minimização dos resíduos (SQE)

processamento da rotina de cálculo da RGP é executado. A Figura 4 exibe, como exemplo de ilustração, a seleção circular das unidades geográficas próximas à unidade em análise (\diamond) estabelecida para o valor da distância $D_{influencia} = 4$ km. Esta seleção, além de diminuir o esforço computacional do processamento, indica quais unidades geográficas apresentarão influência na calibração da reta de regressão da unidade geográfica em análise (\diamond), atribuindo pesos unitários conforme Equação 9 para estas unidades no caso de escolha da ponderação constante.

Já na ponderação gaussiana, estes pesos serão tão mais próximos do valor unitário quanto mais próximos estiverem da unidade geográfica em análise (\diamond). É importante lembrar que o ponto em análise é excluído da calibração de sua própria reta de regressão, conforme mencionado no item 2.2. Estes pesos são então armazenados na matriz W_i (ver Equação 7), com seus coeficientes locais sendo estimados de acordo com a Equação 6. Este processo é então repetido para todas as unidades geográficas do escopo espacial do fenômeno analisado.

O código RGP realiza o processamento anteriormente descrito, desde o valor do parâmetro distância

de influência inicial ($D_{influencia, i}$) até final ($D_{influencia, f}$), com passo de iteração p . Este último *loop* realiza uma pesquisa de minimização dos resíduos, seguindo a Equação 12, e ao final é gerado um arquivo de saída com os valores indicados de SQE e $D_{influencia}$ (conforme ilustrado na Figura 5). Com base no valor ótimo de $D_{influencia}$ obtido — que resulta na minimização de SQE — os valores dos coeficientes do modelo RGP são então finalmente obtidos.

3.4. Visualização temática dos resultados da RGP

Devido à implementação em ambiente SIG, os valores dos *outputs* podem ser analisados na forma de mapas temáticos (Figura 6). Esta forma de representação permite avaliar a variação espacial dos coeficientes modelados pela RGP (intercepto e coeficientes angulares), bem como os resultados dos resíduos e valores modelados da variável dependente y . Pode-se inclusive realizar uma inspeção das unidades geográficas que não foram adequadamente modeladas pela RGP, em função da não existência de similaridade de sinais e ordem de grandeza destas unidades.

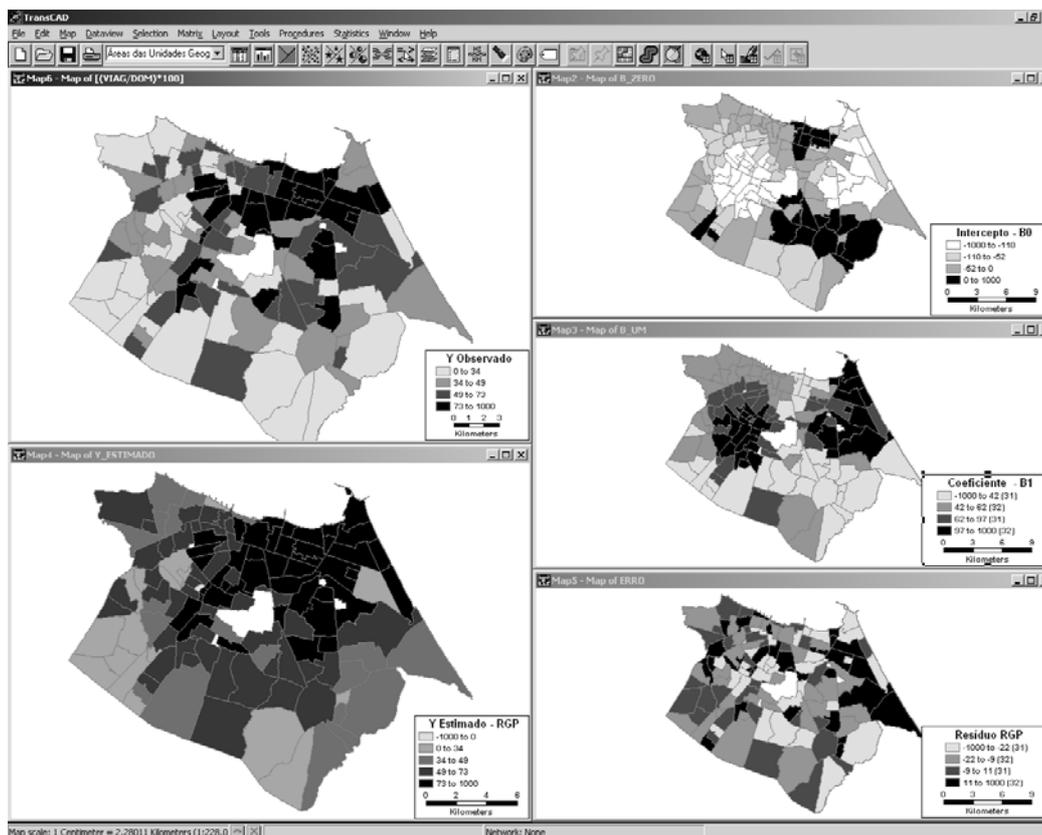


Figura 6: Visualização dos mapas temáticos dos *outputs* da RGP

4. CONSIDERAÇÕES FINAIS

O propósito maior deste artigo foi disseminar o uso de uma ferramenta analítica de estatística espacial, a regressão linear geograficamente ponderada, na compreensão e previsão de fenômenos relacionados aos transportes urbanos e regionais, apresentando uma implementação computacional em ambiente SIG. A abordagem teórica objetivou facilitar a compreensão das premissas básicas e da formulação matemática do modelo RGP, dada a escassa bibliografia existente sobre o assunto. Da mesma forma, na descrição da rotina computacional, buscou-se contemplar ao máximo seus aspectos teóricos, destacando os dados de entrada necessários e os procedimentos de calibração e análise dos resultados do modelo.

Algumas características de sistemas SIG, como tratamento de banco de dados georeferenciados e visualização gráfica, fornecem ferramentas poderosas de análise espacial, tornando tais sistemas adequados para a resolução de uma classe ampla de problemas. Além disso, alguns sistemas disponibilizam plataformas de desenvolvimento, de modo que possam ser ainda mais adaptáveis, abrangendo mais problemas através da implementação de ferramentas de análise ou mesmo solução. Espera-se que a discussão apresentada neste artigo tenha contribuído para demonstrar a adequabilidade, a adaptabilidade e a conveniência do ambiente SIG no desenvolvimento de tais ferramentas.

Por meio da abordagem compreensiva dada à implementação, procurou-se ainda estabelecer conceitos e delinear melhor um paradigma para projetos de algoritmos que favoreça a familiarização com a plataforma e o desenvolvimento de novas ferramentas e aplicações. Neste ponto, a particularização da implementação para uma plataforma específica (GISDK) buscou exemplificar os detalhes de programação, reforçando ainda mais este objetivo. De um modo geral, as técnicas de estatística espacial, como métodos bastante aplicados de regressão de *kernel* e *kriging*, encontram na plataforma SIG um ambiente adequado de desenvolvimento e aplicação, sendo bastante desejável que tais técnicas sejam portadas e adaptadas sob a luz do paradigma SIG. Aplicações mais complexas podem ser melhor tratadas através da aplicação conjunta de ferramentas adicionais de análise, como *clustering*, e algoritmos de otimização.

É importante lembrar que a implementação do modelo RGP descrita neste artigo, apesar de ser suficiente para a obtenção dos coeficientes locais das equações de regressão, deve ser encarada como um passo inicial, e não exaustivo, do assunto, visando a instigar a comunidade técnico-científica brasileira de Transportes para o emprego de modelos de regressão espacial e, de modo geral, de ferramentas de análise espacial exploratória e confirmatória no planejamento e operação de sistemas de transportes, como forma de avançar na utilização da Análise Espacial, fornecendo aos modelos um grau maior de confiabilidade e res-

paldo técnico e científico.

Com relação ao modelo RGP, recomenda-se como prioridade em futuras implementações o desenvolvimento das estatísticas locais dos coeficientes de correlação e determinação, como forma de definir mais claramente o nível de abrangência da modelagem geponderada. Por fim, recomenda-se também a calibração de modelos RGP em paralelo com modelos convencionais (não-espaciais), buscando ampliar a compreensão das premissas e das características intrínsecas dos modelos de regressão espacial, avaliando melhor seu potencial de aplicação.

AGRADECIMENTOS

Ao CNPq pela concessão de bolsas de Produtividade e de Iniciação Científica essenciais para a realização desta pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

- Brusdon, C.; Fotheringham, A. S.; Charlton, M.E. (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4): 281-298.
- Caliper (2002a) *TransCAD - Transportation workstation software, user's guide, Version 4.5 for Windows*. Caliper Corporation, Newton, USA.
- Caliper (2002b) *GISDK - Geographic information system developer's kit*. Caliper Corporation, Newton, USA.
- Câmara, G.; Carvalho, M.S.; Cruz, O. G.; Correa, V. (2002) *Análise espacial de áreas*. In: Fuks, S.D.; Carvalho, M.S.; Câmara, G; Monteiro, A.M.V. (Eds) *Análise espacial de dados geográficos - Divisão de Processamento de Imagens*, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brasil.
- Diggle, P.J.; Tawn, A.; Moyeed, R. A. (1998) Model-based geostatistics. *Applied Statistics*, 47(3): 299-350.
- Fotheringham, A.S.; Charlton, M.; Brusdon, C. (1997) *Recent developments in spatial analysis*, Chapter measuring spatial variations in relationships with geographically weighted regression, pages 60-82. Springer, New York, EUA.
- Fotheringham, A. S. e Brunson, C. e Charlton, M. (2000) *Quantitative geography - Perspectives on spatial data analysis*. Ed. Sage, London.
- Henrique, C.S.; Loureiro, C.F.G.; Cavalcante, R.A. (2005) Caracterização espacial da mobilidade dos usuários cativos do sistema integrado de transportes de Fortaleza. In: *Transporte em Transformação IX - Trabalhos Vencedores do Prêmio CNT Produção Acadêmica 2004*. CNT - Confederação Nacional do Transporte/ANPET - Associação Nacional de Pesquisa e Ensino em Transportes, Brasília, DF, p.61-79.
- Lopes, S. B. (2005) *Efeitos da dependência espacial em modelos de previsão de demanda por transporte*. Dissertação de Mestrado, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, 137 fl.
- Loureiro, C.F.G.; Silva H. N.; Carvalho, L. E. X. (2006) Metodologia de análise de regressão geograficamente ponderada aplicada ao fenômeno das viagens intermunicipais. *Anais do XX Congresso de Pesquisa e Ensino em Transportes*. ANPET - Associação Nacional de Pesquisa e Ensino em Transportes, Brasília, DF, 479-491.
- Silva, A.R. (2006) *Avaliação de modelos de regressão espacial para análise de cenários do transporte rodoviário de carga*. Dissertação de Mestrado. Faculdade de Tecnologia, Universidade de Brasília, DF, 122p.