

METODOLOGIA DE ANÁLISE DE REGRESSÃO GEOGRAFICAMENTE PONDERADA APLICADA AO FENÔMENO DAS VIAGENS INTERMUNICIPAIS

Carlos Felipe Grangeiro Loureiro

Henrique Nogueira Silva

Programa de Mestrado em Engenharia de Transportes - PETRAN

Universidade Federal do Ceará – UFC

Luís Eduardo Ximenes Carvalho

Brown University

RESUMO

Os modelos de regressão linear simples e múltipla têm sido largamente utilizados em estudos de planejamento de transportes, especialmente visando à previsão da demanda na etapa de geração de viagens. Contudo, quando existe autocorrelação espacial da variável dependente, ou mesmo das variáveis explicativas, a análise dos fenômenos de transporte demanda a incorporação do atributo espacial. Este artigo tem, portanto, o objetivo de descrever uma metodologia de calibração de modelos de regressão linear geograficamente ponderada, tendo como exemplo de aplicação a análise da geração de viagens radiais no sistema de transporte rodoviário intermunicipal do Estado do Ceará.

ABSTRACT

Simple and multiple linear regression models have been largely used in transportation planning studies, specially aiming at forecasting demand during the trip generation stage. However, when there is spatial autocorrelation for the dependent variable, or even for the explicative variables, the analysis of transportation phenomenon requires the consideration of the spatial attribute. This paper has, therefore, the objective of describing a methodology for calibrating geographically weighted linear regression models, presenting as a sample application the analysis of radial trip generation in the intercity public transit system of the State of Ceara.

1. INTRODUÇÃO

Os modelos de regressão linear simples e múltipla têm sido largamente utilizados em estudos de planejamento de transportes, especialmente visando à previsão da demanda na etapa de geração de viagens. Estes modelos são normalmente empregados de forma agregada para prever a quantidade de viagens produzidas ou atraídas em uma dada zona, tendo como base um conjunto de variáveis explicativas relacionadas à sua demografia e atividade socioeconômica. Contudo, quando existe autocorrelação espacial da variável dependente, ou mesmo das variáveis explicativas, a análise dos fenômenos de transporte demanda a incorporação do atributo espacial (Lopes, 2005). Este fato é corroborado quando se considera que a maioria das atividades de transporte de passageiros e carga está relacionada com atributos eminentemente geográficos, tais como: adensamentos residenciais e de atividades socioeconômicas, proximidades entre zonas, cobertura espacial da rede de transporte, impedância de viagem na malha rodoviária, dentre outros.

Desta forma, a aplicação de modelos tradicionais de regressão linear, sem a incorporação de atributos geográficos, pode conduzir à superestimação da significância dos parâmetros dos modelos, além de evidenciar associações espúrias (Câmara *et al.*, 2002). Para evitar problemas dessa natureza na modelagem da demanda por transportes, a incorporação do caráter espacial na calibração de modelos de regressão linear pode ser efetuada por meio da técnica de ponderação geográfica (Fotheringham *et al.*, 2000). Em modelos de geração de viagens geograficamente ponderados a premissa é de que as observações mais próximas à zona i possuem maior influência na estimativa dos coeficientes da regressão que pontos mais afastados da localização de i . Cabe destacar que a modelagem geponderada tem necessariamente um efeito local, na medida que assume a hipótese de não estacionariedade

espacial da variável dependente, conforme explicado por Carvalho *et al.* (2006). Seu grande diferencial em termos de poder explicativo é produzir, em vez de uma única relação de dependência válida para toda a área de estudo (modelo global), relações específicas para cada região (modelos locais) em que ocorrem influências locais motivadas pela autocorrelação espacial do fenômeno analisado.

Este artigo tem, portanto, o objetivo de descrever uma metodologia de calibração de modelos de regressão linear geograficamente ponderada, 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 (STRIP-CE).

2. METODOLOGIA DE ANÁLISE DE REGRESSÃO GEOPONDERADA

A metodologia de análise de regressão geograficamente ponderada (RGP) apresentada neste artigo busca trazer contribuições ao processo de modelagem tradicional baseado na calibração de funções de regressão linear não-espaciais, cujo entendimento acredita-se já se encontrar em nível bem consolidado por parte dos analistas do sistema de transportes. Com essa proposta metodológica, representada no fluxograma da Figura 1, pretende-se disseminar o uso de uma potente ferramenta de análise confirmatória que evidencie os aspectos positivos e negativos da consideração do atributo espacial na compreensão e explicação dos fenômenos inerentes ao planejamento de transporte. A descrição detalhada das etapas desta metodologia é apresentada a seguir.

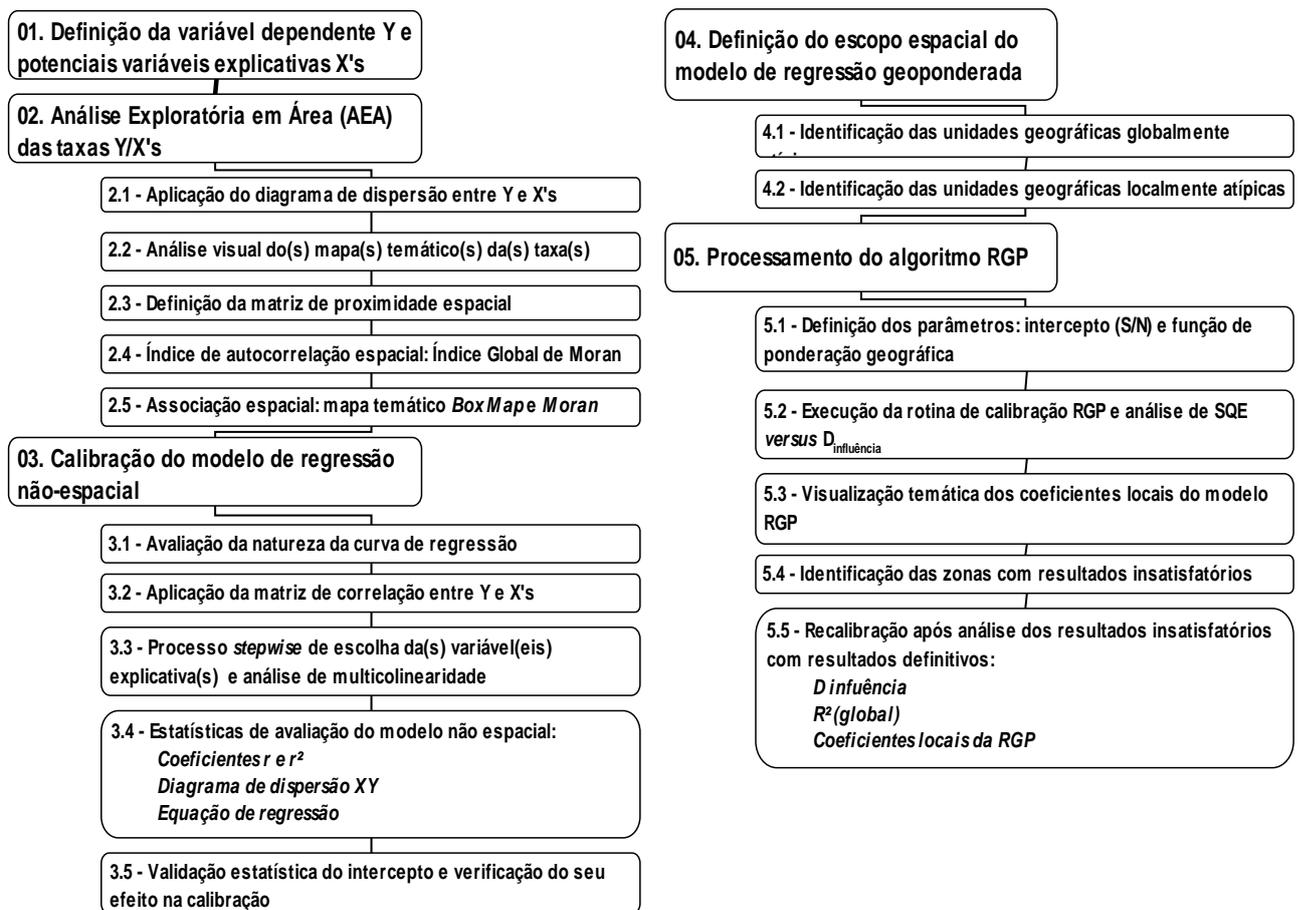


Figura 1: Metodologia de análise de regressão geograficamente ponderada.

2.1. Definição da variável dependente Y e potenciais variáveis explicativas X's

Inicialmente, deve-se definir de forma clara a variável dependente Y que irá representar o fenômeno em análise, bem como especificar um grupo de variáveis potencialmente capazes de explicar o comportamento da variável dependente. Neste grupo de candidatas a variáveis explicativas é importante estar atento para evitar a inclusão de variáveis X's estatisticamente correlacionadas com Y, mas que não possuem associação fenomenológica com a variável dependente. Dessa forma, pretende-se trabalhar com variáveis explicativas que, além de possuírem elevada correlação com o comportamento da variável Y, apresentem também uma relação de causalidade com o fenômeno a ser explicado.

2.2. Análise Exploratória em Área (AEA) das taxas Y/X's

O objetivo desta etapa é possibilitar ao analista avaliar qualitativamente a necessidade da consideração do atributo espacial na explicação do fenômeno estudado. Neste estágio, considera-se a hipótese de que, em fenômenos eminentemente geográficos, os modelos confirmatórios espaciais traduzem melhor o padrão de comportamento destes fenômenos, ou seja, ajustes de modelos não-espaciais e globais (como é o caso da regressão tradicional) não expressarão bem as relações de interdependência esperadas. A verificação desta hipótese deve ser feita inicialmente a partir dos diagramas de dispersão de Y versus cada X, buscando identificar o tipo de relação de dependência não-espacial entre cada par de variáveis. Caso não se constatem nos diagramas de dispersão fortes relações de dependência, lineares ou não, entre variáveis que se supõe apresentarem relações inerentes de causalidade, deve-se investigar a contribuição que o caráter espacial destas variáveis pode trazer à compreensão do fenômeno em análise.

Esta investigação pode ser realizada com o uso de técnicas de Análise Exploratória em Áreas (AEA), partindo-se inicialmente da geração e análise de mapas temáticos das taxas Y/X_i e complementando com mapas de médias móveis e de indicadores locais de associação espacial (p.ex., índice local de Moran). Para tanto, deve-se definir uma matriz de vizinhança ou proximidade espacial para modelar a relação topológica entre as unidades geográficas em análise, sendo esta matriz de primeira ordem (unidades adjacentes) ou de ordem superior (Anselin, 2005). Em geral, a matriz de proximidade de primeira ordem é suficiente para a AEA, com a matriz de proximidade de segunda ordem (ou superior) sendo utilizada para um exame de tendência dos resultados da AEA.

No passo seguinte da AEA, o cálculo do Índice Global de Moran (I) servirá como indicador para quantificar o nível de autocorrelação espacial da variável dependente, assim como das potenciais variáveis explicativas. Pode-se avaliar nesta etapa uma relação direta, expressa pelo Índice Global de Moran, entre o nível de autocorrelação espacial e a expectativa de incremento no poder de explicação do modelo de regressão devido à inclusão do atributo espacial. Por fim, os mapas de associação espacial local (*Box Map* e *Moran Map*) indicarão as unidades geográficas que possuem comportamento semelhante (associação positiva) ou inverso (associação negativa), estabelecendo aglomerados uniformes ou bolsões de estacionariedade. Pode-se afirmar que o incremento na explicação de Y promovido pelo modelo RGP sobre o modelo tradicional será maior à medida que a quantidade de unidades geográficas com associação espacial positiva crescer. Mesmo a AEA apontando fortes indícios de dependência espacial no fenômeno analisado, recomenda-se a calibração de modelos tradicionais de regressão linear, com ou sem transformação de variáveis, antes de se partir para a análise geponderada.

2.3. Calibração do modelo de regressão não-espacial

Esta etapa deve ser entendida como uma fase de investigação e entendimento do processo de calibração de modelos de regressão, servindo de suporte para o posterior desenvolvimento do modelo espacial (RGP), uma vez que espera-se que este modelo proporcione um incremento no poder de explicação do modelo tradicional. Como se supõe que a compreensão do processo de calibração de modelos de regressão não-espacial já esteja suficientemente consolidada na comunidade técnico-científica de Transportes, apresenta-se a seguir apenas uma síntese dos principais aspectos a serem considerados neste processo, baseada no trabalho de Neter *et al.* (1990).

Como primeiro passo, recomenda-se a definição da função (linear ou não) que melhor expressa a relação de dependência entre Y e cada X, com base nos diagramas de dispersão e nos coeficientes de correlação linear de Pearson. É importante ressaltar que o uso de um modelo não-linear é justificado apenas quando o seu poder de explicação for significativamente superior ao do modelo linear ($r^2_{\text{não-linear}} \gg r^2_{\text{linear}}$). Uma vez estabelecida a forma da relação de dependência, constrói-se a matriz de correlação linear (com transformação de variáveis, quando recomendado). Tal matriz indicará um *ranking* de variáveis X's que entrarão no processo *stepwise* de seleção das variáveis explicativas que melhor se adequam ao modelo de regressão não-espacial. É importante ressaltar ainda a necessidade da avaliação da homocedasticidade e da multicolinearidade entre as variáveis explicativas, já que o processo *stepwise* estabelece a minimização dos resíduos como critério único. Uma vez definidos o caráter das relações de dependência, assim como quantas e quais variáveis explicativas farão parte do modelo de regressão não-espacial, deve-se calibrar tal modelo, não esquecendo de avaliar a significância estatística do intercepto e o seu efeito na significância dos coeficientes angulares das variáveis selecionadas.

2.4. Definição do escopo espacial do modelo de regressão geograficamente ponderada

Nesta fase, são identificadas as unidades geográficas (UG's) atípicas cujo comportamento deve ser compreendido e justificado separadamente sob o ponto de vista diretamente associado ao fenômeno em análise (p.ex., uso do solo ou socioeconomia). Esta atipicidade poderá ser de uma intensidade tal que resultará na exclusão destas UG's do escopo espacial do modelo RGP. Caso não sejam excluídas, a ponderação geográfica proposta pela RGP poderá resultar em coeficientes da curva de regressão destoantes da realidade nestas UG's, uma vez que a hipótese de interação *unidade geográfica – vizinhança*, verificada para a maioria das unidades geográficas, pode não ser válida para estas UG's atípicas. Existirão basicamente dois tipos de UG's atípicas:

- (a) unidades geográficas globalmente atípicas: são UG's cuja posição no diagrama de dispersão XY está essencialmente distante do padrão da curva de regressão global não-espacial. Pode-se adotar como critério para estabelecer unidades globalmente atípicas aquelas com resíduos superiores a 3 desvios padrões, positivos ou negativos. Vale destacar que, no caso destas UG's apresentarem associação local positiva, pode-se desejar incluí-las no modelo RGP, se este for calibrado com o intuito de realizar previsão futura de um fenômeno no qual a tendência temporal é desta UG influenciar seus vizinhos;
- (b) unidades geográficas localmente atípicas: são UG's identificadas no *Box Map* com associação espacial negativa, provavelmente com significância destacada no *Moran Map*, devidamente justificadas pelos aspectos inerentes ao fenômeno analisado. Sua exclusão do modelo evitará a representação de uma influência sobre os seus vizinhos (e vice-versa) que não se verifica no fenômeno real, atualmente ou no futuro.

2.5. Processamento do algoritmo RGP

Após a definição do escopo espacial de modelagem da regressão geograficamente ponderada, inicia-se o processamento do algoritmo RGP. Este algoritmo foi implementado por meio de uma rotina computacional, escrita em linguagem GISDK, no pacote TransCAD (Caliper, 2002) de Sistema de Informações Geográficas aplicado a Transportes (SIG-T), conforme descrito em Carvalho *et al.* (2006). Acredita-se ser esta a primeira implementação de um algoritmo de RGP em ambiente SIG. Essa rotina processa os dados de entrada para obtenção dos coeficientes da curva de regressão local das UG's e possui a flexibilidade de exclusão do intercepto e de escolha do tipo de ponderação geográfica (constante ou gaussiana).

Após a execução desta rotina, é recomendável a visualização dos mapas temáticos dos coeficientes angulares calibrados no modelo de regressão. Ao analisar a distribuição espacial dos valores destes coeficientes, será possível identificar algumas UG's em que o algoritmo RGP apresentou resultados pouco satisfatórios, ainda que estes não tenham sido identificados na análise exploratória. Por fim, pode-se adotar algum critério de exclusão destas UG's e processar a rotina RGP novamente com a obtenção dos valores definitivos dos seguintes parâmetros: (i) distância de influência; (ii) coeficiente de determinação (global); e (iii) coeficientes locais da curva de regressão.

3. APLICAÇÃO DA RGP NO FENÔMENO DAS VIAGENS INTERMUNICIPAIS

De modo a ilustrar o processo de calibração de modelos de regressão geoponderada, a metodologia proposta foi aplicada na análise da geração de viagens interurbanas radiais (com destino a Fortaleza) no Sistema de Transporte Rodoviário Intermunicipal do Estado do Ceará (STRIP-CE). A escolha deste fenômeno para exemplo de aplicação da RGP se deu em função da hipótese de forte dependência socioeconômica com a capital dos municípios do interior do Estado (excluindo aqueles que integram a Região Metropolitana de Fortaleza) – considerados como unidades geográficas nesta análise. Os dados socioeconômicos e de viagens realizadas nos modos regulares (ônibus) e clandestinos (vans, caminhões e caminhonetes) foram obtidos a partir dos estudos de elaboração do Plano Diretor e Operacional do Transporte Intermunicipal de Passageiros do Estado do Ceará (PDOTIP-CE) (ASTEF, 2005), tendo 1997 como ano-base.

3.1. Definição das variáveis Y e X's

Considerando a primeira etapa da metodologia proposta, a variável dependente (Y) foi definida como o número de viagens radiais geradas (produzidas e atraídas) por um determinado município, somando os deslocamentos realizados em 1997, tanto no serviço regular por ônibus, como nos veículos de transporte clandestino. Já o conjunto das potenciais variáveis explicativas (X's) abrangeu variáveis de caráter socioeconômico e do sistema de transporte, todas agregadas no nível de município. Dessa forma, foram consideradas variáveis de população (total, por faixa etária, urbana e rural) e de número de domicílios, além do PIB, transferência do governo estadual, do governo federal e consumo de energia, ao passo que a principal variável do sistema de transporte utilizada foi a distância de cada município a Fortaleza pela rede de transporte público.

3.2. AEA da taxa de viagens radiais

Após a elaboração dos diagramas de dispersão XY para cada uma das potenciais variáveis explicativas, verificou-se que a variável *População de 18 a 64 anos* estava entre aquelas que apresentavam uma relação de dependência linear mais forte com a variável Y, embora

apresentando elevada correlação linear com a maioria das demais variáveis X's. Acredita-se que esta variável explicativa tenha uma relação de causalidade direta com o fenômeno analisado, já que a população usuária do transporte público intermunicipal normalmente não inclui crianças e adolescentes em idade escolar (atendidas pela rede municipal de ensino), nem idosos (estes tendem a apresentar uma taxa de mobilidade intermunicipal bastante reduzida). Portanto, esta variável explicativa foi escolhida para ilustrar o processo de análise exploratória em áreas da variável *Taxa de Viagens Radiais por População de 18 a 64 anos*, a fim de investigar a relação de dependência geográfica existente no fenômeno das viagens radiais. O mapa temático desta taxa Y/X está apresentado na Figura 2, observando-se uma tendência de elevação da taxa à medida que estes municípios se aproximam da Região Metropolitana de Fortaleza – RMF (excluída do mapa). Este comportamento reforça o caráter espacial do fenômeno, além de sugerir a consideração da variável locacional *Distância do Município a Fortaleza*, como forma de melhorar o poder de explicação do modelo de regressão não-espacial.

Em seguida, foi calculado o Índice Global de Moran com o uso do pacote computacional GeoDA (SAL, 2004). O valor obtido de $I = 0,446$, estatisticamente significativo (valor-p = 0,001), confirma a hipótese de que este fenômeno possui uma forte dependência espacial, indicando que a ponderação geográfica provavelmente resultará em acréscimo significativo no poder de explicação do modelo de regressão. Porém, a simples análise do Índice Global de Moran não explicita o comportamento do fenômeno nas diversas regiões do Estado. Como dito anteriormente, o tipo de associação espacial que os municípios tem com sua vizinhança representa uma informação relevante para a calibração do modelo espacial, principalmente em modelos locais como é o caso da RGP. A Figura 3 exhibe o *Box Map* da taxa de viagens radiais, tomando por base a matriz de proximidade espacial de primeira ordem. A associação geográfica dos municípios pode ser classificada em dois grandes grupos, de acordo com os quadrantes do Diagrama de Espalhamento de Moran:

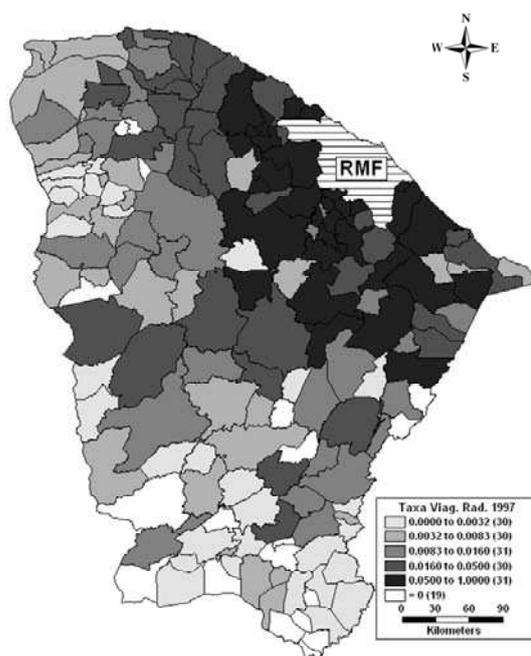


Figura 2: Taxa de viagens radiais em 1997.

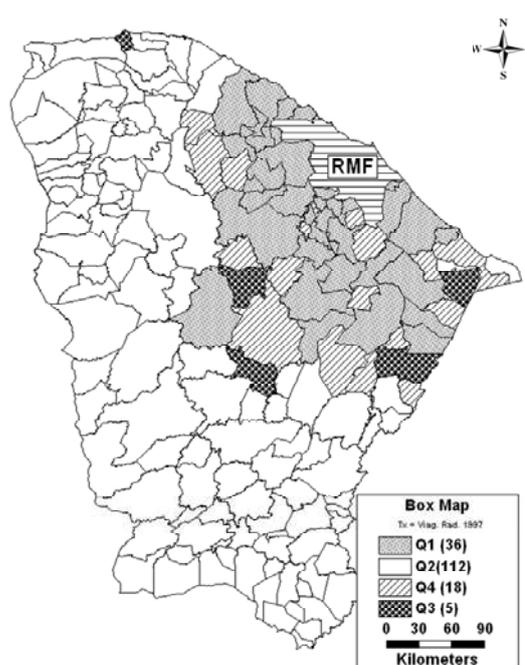


Figura 3: Associação espacial da taxa de viagens radiais em 1997 – *Box Map*.

Grupo 1: Municípios com associação espacial semelhante a sua vizinhança (positiva)

- a) Quadrante 1 – taxa do município e dos seus vizinhos acima da média – pode-se perceber que os municípios neste quadrante se agrupam em torno da RMF, formando assim um “grande cinturão” de intensa comunicação e dependência socioeconômica com Fortaleza. Nestes municípios, o modelo RGP irá estimar curvas de regressão com maiores coeficientes angulares, modelando de forma mais realista suas viagens radiais do que o modelo de regressão global (regressão linear simples não-espacial - RLS).
- b) Quadrante 2 – taxa do município e dos seus vizinhos abaixo da média – municípios com comportamento inverso ao do Quadrante 1. Todos os municípios neste quadrante estão a uma distância superior a 120 km de Fortaleza e provavelmente trocam mais viagens com municípios pólos (concentradores de atividades socioeconômicas) mais próximos, resultando numa fraca relação de dependência com Fortaleza. Nesta classe também haverá ganhos com a ponderação geográfica da RGP, haja vista que este modelo espacial irá descrever curvas de regressão menos inclinadas e mais realistas que a RLS.

Grupo 2: Municípios com associação espacial inversa à da sua vizinhança (negativa)

- c) Quadrante 3 – taxa do município acima da média e taxas dos vizinhos abaixo da média - municípios localmente atípicos em razão de sua elevada taxa de viagens radiais, inseridos em uma região formada por municípios com taxas reduzidas. A consideração de um município desta classe no modelo RGP não é recomendada, pois o efeito da ponderação geográfica resultaria em: (i) considerável subestimação do coeficiente angular da reta de regressão deste município, contrariando assim a realidade observada no mesmo; e (ii) superestimação do coeficiente angular da reta de regressão dos municípios vizinhos. Municípios neste quadrante representam normalmente pólos regionais que exercem forte influência socioeconômica sobre seus vizinhos, mantendo, porém, uma relação relativamente forte de dependência com Fortaleza.
- d) Quadrante 4 – taxa do município abaixo da média e taxas dos vizinhos acima da média - tratam-se de municípios que apesar de estarem situados no “grande cinturão” em torno da RMF, apresentam taxas relativamente baixas quando comparadas às dos seus vizinhos. A maioria destes municípios apresenta baixos níveis de atividade econômica, podendo-se afirmar que essas reduzidas taxas de viagens radiais são consequência de um estágio de desenvolvimento ainda não consolidado. Em se constatando esta situação, pode-se lançar a hipótese que estas UG's tenderão no futuro a apresentar níveis cada vez maiores de intercâmbio de viagens com Fortaleza, em função de sua proximidade. Para estes municípios haverá uma considerável superestimação no coeficiente angular de sua curva local de regressão, o que pode ser desejável com base na hipótese de “retardo de desenvolvimento econômico e de dependência em relação à Capital”, caso o modelo de regressão seja utilizado para fins de previsão futura de demanda.

3.3. Calibração do modelo de regressão não-espacial para viagens radiais

Como passo inicial da calibração do modelo de regressão tradicional, foram gerados, utilizando o pacote computacional SPSS (2001), os diagramas de dispersão XY buscando-se ajustar curvas globais não-espaciais de regressão de diferentes naturezas. As Figuras 4 e 5 apresentam o diagrama da variável *População de 18 a 64 anos*, verificando-se neste caso que as observações não se ajustam perfeitamente nem ao padrão linear – principalmente para municípios com população entre 20 e 40 mil habitantes, nos quais as viagens radiais tendem a

apresentar valores bastante dispersos – e muito menos aos padrões não-lineares testados. Procedimento semelhante foi realizado para a análise de cada uma das variáveis candidatas a explicarem Y, chegando-se à conclusão que em todas elas a função linear representaria melhor sua relação de dependência com a variável *Viagens Radiais*.

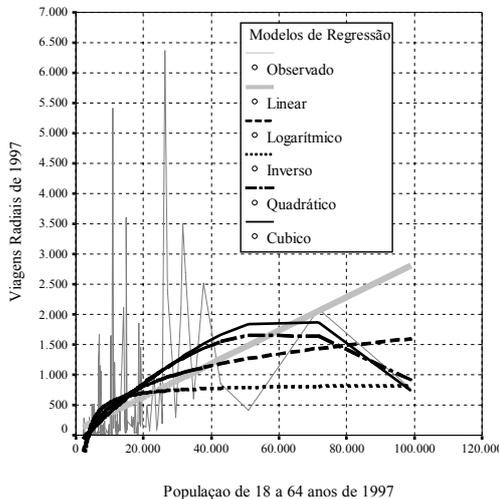


Figura 4: Diagrama de dispersão XY e curvas globais não-espaciais de regressão.

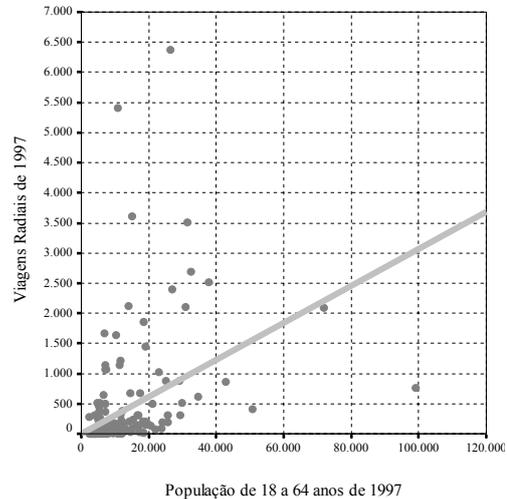


Figura 5: Diagrama de dispersão XY das viagens radiais 1997 e reta de regressão global.

Uma vez definido que o modelo empregado deveria ser do tipo linear, construiu-se a matriz de correlação linear, sem transformação de variáveis, apresentada parcialmente na Tabela 1. Verifica-se na matriz que as sete variáveis explicativas consideradas, apesar de serem aquelas que melhor se ajustaram à variável Y dentre um conjunto bem maior de variáveis demográficas e socioeconômicas, não apresentaram um valor considerável de correlação linear ($r > 0,5$) com a variável Y, com todas apresentando coeficientes muito próximos, variando entre 0,26 e 0,47. O processo *stepwise* de seleção de variáveis (SPSS, 2001), com probabilidade de 5% de entrada e 10% de saída para a estatística F, indicou duas variáveis X's para consideração no modelo: *Distância a Fortaleza* e *População Total*. Este modelo não apresentou problemas de multicolinearidade, tendo um coeficiente de determinação (r^2) igual a 37%.

Tabela 1: Matriz de correlação linear das variáveis analisadas.

		Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
		<i>Viagens Radiais</i>	Pop. Total	Pop. Urbana	Pop. Rural	Pop. 18 a 64 anos	Cons. Energ. Públ.	PIB	Dist. Fortaleza
Y	<i>Viagens Radiais</i>	1,000	-	-	-	-	-	-	-
X ₁	Pop. Total	0,371	1,000	-	-	-	-	-	-
X ₂	Pop. Urbana	0,345	0,952	1,000	-	-	-	-	-
X ₃	Pop. Rural	0,260	0,645	0,379	1,000	-	-	-	-
X ₄	Pop. 18 a 64 anos	0,371	0,998	0,961	0,615	1,000	-	-	-
X ₅	Cons. Energ. Públ.	0,398	0,817	0,754	0,584	0,821	1,000	-	-
X ₆	PIB	0,350	0,833	0,844	0,411	0,839	0,703	1,000	-
X ₇	Dist. Fortaleza	0,468	0,057	0,074	0,013	0,060	0,045	0,003	1,000

Ao aplicar-se o modelo proposto pelo procedimento *stepwise*, verificou-se que este modelo de regressão linear múltipla resultava em 37 municípios (21% do total) com número negativo de viagens radiais previstas e, caso se forçasse seu intercepto a zero, seu poder de explicação seria drasticamente reduzido. Esta constatação levou à consideração de modelos de regressão simples, tendo como candidata mais adequada à variável explicativa a *Distância a Fortaleza*, já que possui o maior coeficiente de correlação linear com Y. Porém, este modelo foi descartado em função de não contemplar a relação das viagens radiais com o nível de atividade socioeconômica do município, resultando em um modelo estático.

Estas considerações levaram à escolha da variável *População de 18 a 64 anos* como variável X mais adequada, inclusive por acreditar-se que esta variável possui uma relação de causalidade com as viagens radiais mais evidente que a variável *População Total*, conforme justificado anteriormente. Após a definição da variável explicativa, foi realizada uma calibração inicial da RLS e verificou-se que o intercepto da equação de regressão não apresentava valor estatisticamente significativo. Procedeu-se, então, uma segunda calibração de RLS em que forçou-se o intercepto ao valor nulo, resultando em aumento do poder de explicação do modelo. A representação gráfica da equação da RLS (Figura 5) deixa clara a existência de cinco municípios globalmente atípicos: quatro *outliers* positivos e um negativo. Pode-se perceber também que, para todas as faixas de população, o modelo global não se ajusta bem, com todas elas apresentando municípios com resíduos elevados, especialmente aquelas com população até 40 mil habitantes, ou superior a 80 mil, na faixa etária de 18 a 64 anos.

3.4. Definição do escopo espacial da RGP para viagens radiais

Para os municípios globalmente atípicos destacados no item anterior, pode-se verificar que, caso se optasse pelo modelo de regressão não-espacial (RLS), estes deveriam ser excluídos do modelo de previsão a fim de garantir menor dispersão dos dados, resultando em melhoras no poder de explicação do modelo global. Porém uma análise local (*Box Map* da Figura 3) evidencia que todos os cinco municípios globalmente atípicos não são considerados como localmente atípicos; em outras palavras, a região geográfica em que estes se situam segue, em menor proporção, o seu padrão. Portanto, optou-se por mantê-los no modelo RGP para avaliar a magnitude do seu efeito de influência sobre as taxas previstas para os seus vizinhos. Vale destacar que esta decisão deve ser precedida de uma análise socioeconômica da região em questão. Já os municípios localmente atípicos excluídos do modelo RGP são aqueles pertencentes ao Quadrante 3 pelos motivos mencionados no item 3.2.

3.5. Processamento do algoritmo RGP para viagens radiais

Após a definição do escopo geográfico de modelagem da RGP, avaliou-se inicialmente dois parâmetros importantes deste modelo espacial: (i) efeito do intercepto – pode-se excluí-lo ou não na modelagem; (ii) curva de ponderação geográfica – do tipo constante ou gaussiana. O terceiro parâmetro é avaliado pela curva de calibração da *Soma do Quadrado dos Erros x Distância de Influência* ($SQE \times D_{\text{influência}}$) para as quatro combinações dos parâmetros (i) e (ii). É importante observar que apesar da RLS não ser um modelo de regressão espacial, foi calibrada a fim de podermos avaliar o nível de incremento no poder de explicação da variável X devido ao efeito da ponderação geográfica.

O gráfico da Figura 6(a) permite tirar as seguintes conclusões sobre a regressão geograficamente ponderada:

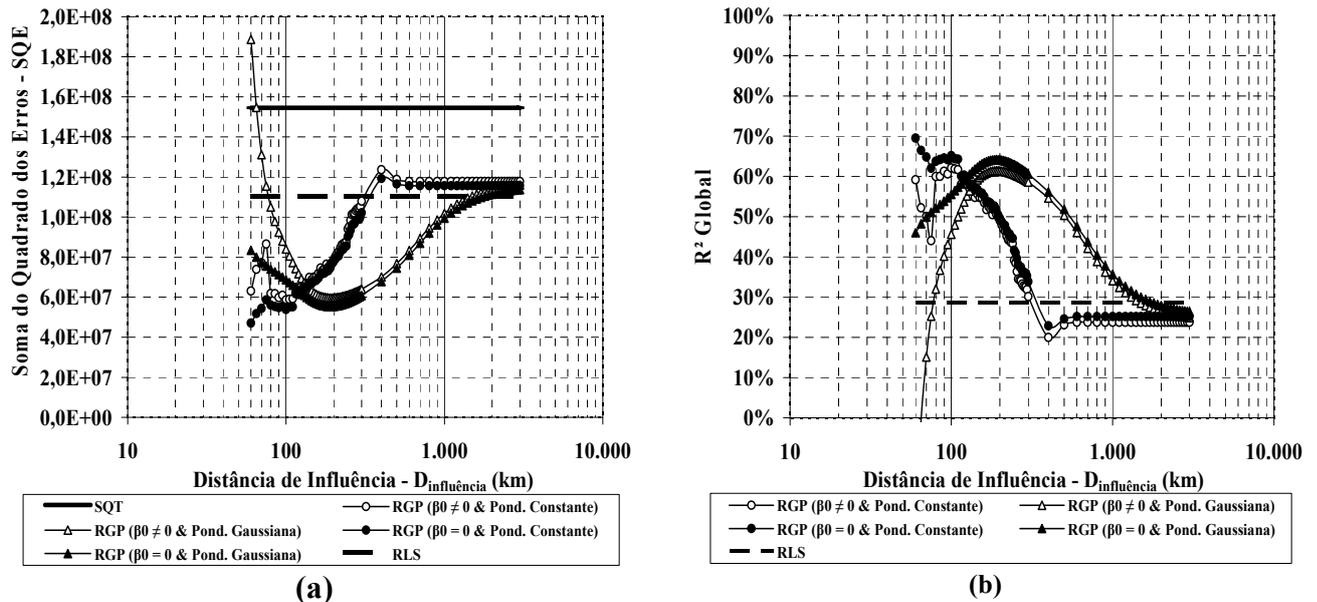


Figura 6: Calibração da RGP para as viagens radiais de 1997.

- Todos os modelos geoponderados apresentaram pico de mínimo com SQE menor que o SQE do modelo RLS, confirmando a hipótese do item 2.2 da metodologia proposta que a consideração espacial em fenômenos geográficos resulta em modelos confirmatórios mais adequados;
- A estatística SQE dos modelos geoponderados tende à estatística SQE da RLS para valores de $D_{\text{influência}}$ suficientemente grandes. Esta observação pode ser facilmente justificada considerando que para distâncias de influência elevadas, acima de 2.000 km para este caso específico, a função de ponderação tende a valores unitários e, conseqüentemente, o modelo RGP tende ao modelo RLS;
- Os modelos geograficamente ponderados com intercepto forçado ao valor nulo ($\beta_0 = 0$) apresentaram uma redução maior no SQE que os respectivos modelos com intercepto não-nulo ($\beta_0 \neq 0$). Esta constatação também foi verificada para o caso do modelo RLS (ver item 3.3), sugerindo que a não significância estatística do intercepto no modelo RLS serve como indicador da não consideração do intercepto na RGP;
- Nos modelos com ponderação constante, os valores de SQE apresentaram mínimo absoluto para distância de influência em torno de 60 km, nesta faixa, porém, a quantidade de observações utilizadas para a calibração da curva de regressão de um município qualquer é demasiadamente pequena (chegando ao valor de somente duas observações para o município de Independência) e não se constitui um pico robusto de SQE;
- Os modelos com função de ponderação gaussiana apresentaram picos de mínimo mais bem definidos e com uma maior abrangência de observações em relação à ponderação constante ($D_{\text{infl., pond. gaussiana}} \approx 190 \text{ km} > D_{\text{infl., pond. constante}} \approx 100 \text{ km}$). Esta maior quantidade de observações favorece a diminuição da estimativa intervalar dos coeficientes modelados (embora esta estimativa não tenha sido discutida neste artigo). Neste ponto, vale destacar que o parâmetro *Distância de Influência* não possui relação alguma com a *Distância Média de Viagem para Fortaleza*. A primeira tenta mensurar uma zona circular de influência das observações sobre as unidade geográficas e é admitida como um valor constante, enquanto que a segunda avalia o nível de impedância das viagens radiais dos municípios para a capital do Estado.

Conseqüência direta do gráfico da Figura 6(a), o gráfico da Figura 6(b) demonstra que os modelos geograficamente ponderados possuem um poder de explicação consideravelmente superior ao modelo RLS, chegando a atingir níveis de explicação de até 65%, contrastando com o coeficiente de determinação de 29% da RLS.

Para a avaliação local do modelo geograficamente ponderado, foram construídos dois mapas temáticos dos coeficientes angulares da reta de regressão, para comparar os resultados obtidos nos dois tipos de ponderação. Na Figura 7, tem-se o mapa dos coeficientes com a ponderação constante, enquanto a Figura 8 refere-se à ponderação gaussiana. Como se pode observar, independente do tipo de ponderação geográfica escolhida, o modelo RGP estima uma reta de regressão específica para cada município que, em linhas gerais, se assemelha ao comportamento espacial da *Taxa de Viagens Radiais* (ver Figura 2). Nesta aplicação, o município de Iguatu é o exemplo mais evidente do efeito indesejável da inversão de coeficientes angulares entre alguns municípios pólos e sua respectiva vizinhança. Nas Figuras 7 e 8, pode-se perceber que a RGP modelou para Iguatu um coeficiente angular menor que sua vizinhança, contrariando a realidade observada no mapa das taxas da Figura 2. Esta inversão revela-se como a única desvantagem da ponderação geográfica e, após a exclusão de tais UG's, com um posterior tratamento isolado (não espacial), recomenda-se executar novamente a rotina de cálculo RGP com os valores definitivos para (i) *Distância de Influência*, (ii) r^2 (global) e (iii) coeficientes locais da RGP.

Já o mapa da Figura 9 foi elaborado com o único intuito de destacar a forte simplificação da consideração de um único valor para o coeficiente angular da reta de regressão. Comparando as Figuras 8 e 9, verifica-se que apenas uma parcela das UG's do modelo RGP apresentaram valores em torno do coeficiente global da RLS. Vale lembrar ainda que, neste exemplo de aplicação, a consideração do atributo espacial é um fator bem mais determinante no poder de explicação do modelo de regressão que a tentativa de inserir outras variáveis explicativas, ou mesmo testar outros tipos de relações não lineares.

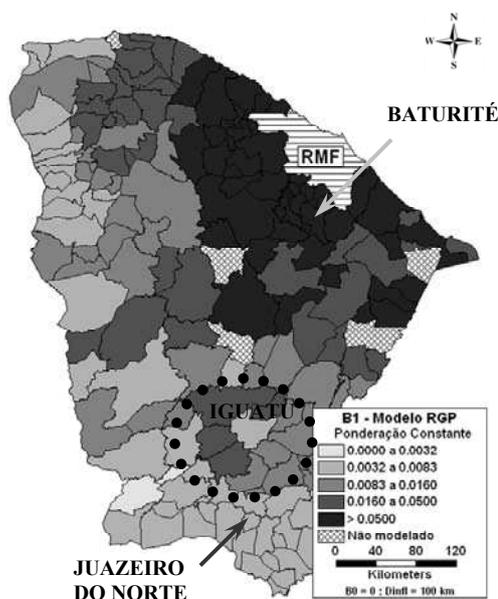


Figura 7: Coeficientes angulares (β_1) da RGP com ponderação constante.

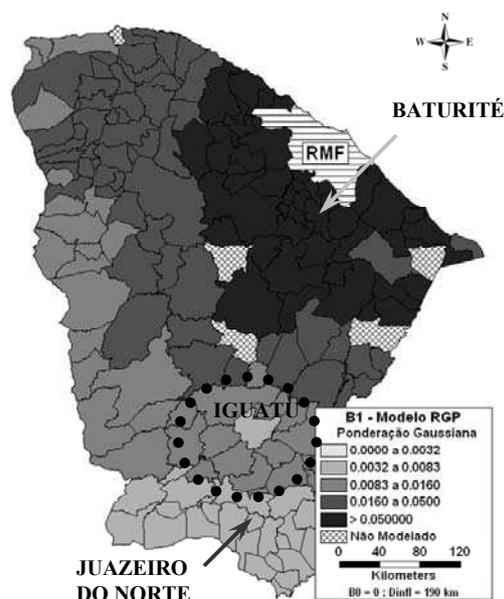


Figura 8: Coeficientes angulares (β_1) da RGP com ponderação gaussiana.

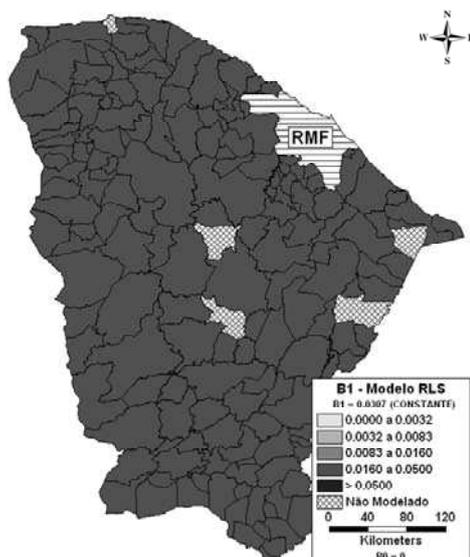


Figura 9: Coeficiente angular (β_1) da RLS.

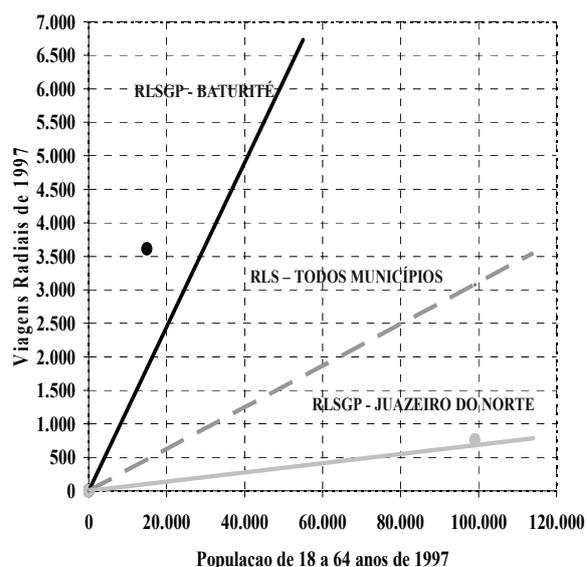


Figura 10: Retas de regressão das viagens radiais: modelo global RLS e modelo local RGP.

A Figura 10 mostra analiticamente o efeito da ponderação geográfica da RGP. Esta ponderação resulta em retas mais inclinadas para municípios mais próximos de Fortaleza (caso de Baturité) e menos inclinadas para municípios distantes da capital (caso de Juazeiro do Norte), resultando em resíduos consideravelmente menores que os da RGP.

4. CONSIDERAÇÕES FINAIS

A ênfase deste artigo no ferramental da Análise Espacial não é um fim em si mesmo, se constituindo apenas numa tentativa de incorporar de uma forma mais eficiente e eficaz estas técnicas quantitativas e qualitativas nos processos de análise dos sistemas de transportes, dado o caráter explicitamente espacial destas análises. Portanto, um dos objetivos secundários do presente trabalho foi o de incentivar a iniciativa de se estudar e utilizar mais intensamente as ferramentas de *análise exploratória* e *análise confirmatória* (modelagem) na análise, planejamento e operação de transportes, como forma de avançar na utilização da Análise Espacial, fornecendo aos modelos um grau maior de confiabilidade e respaldo técnico e científico.

Agradecimentos

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

REFERÊNCIAS BIBLIOGRÁFICAS

- Anselin, L. (2005) – Contiguity-Based Spatial Weights. In: Anselin (ed.) *Exploring Spatial Data with GeoDaTM: A Workbook*. Center for Spatially Integrated Social Science, Santa Barbara, USA.
- ASTEF (2005) *Caracterização Territorial, Socioeconômica, Operacional, Regulatória e Institucional do Sistema de Transporte Rodoviário Intermunicipal de Passageiros do Estado do Ceará*. 2º Relatório Técnico do Plano Diretor Operacional de Transporte Intermunicipal de Passageiros do Estado do Ceará (PDOTIP-CE). Associação Técnico-Científica Eng. Paulo de Frontin, Fortaleza, CE.
- Caliper (2002) *TransCAD - Transportation Workstation Software, User's Guide*, Version 4.5 for Windows. Caliper Corporation, Newton, USA.
- Câmara, G.; Carvalho, M.S.; Cruz, O. G.; Correa, V. (2002) *Análise Espacial de Áreas*, em: *Análise Espacial de Dados Geográficos*, eds. Fuks, S.D.; Carvalho, M.S.; Câmara, G; Monteiro, A.M.V. – Divisão de Processamento de Imagens – Instituto Nacional de Pesquisas Espaciais – São José dos Campos, Brasil.

- Carvalho, L. E. X.; Silva H. N. e Loureiro, C. F. G. (2006) Regressão Geograficamente Ponderada em Ambiente SIG. *Trabalho submetido ao XX Congresso de Ensino e Pesquisa em Transportes*, ANPET, Brasília.
- Fotheringham, A. S. e Brunson, C. e Charlton, M. (2000) *Quantitative Geography – Perspectives on Spatial Data Analysis*. Ed. Sage, London.
- Lopes, S. B. (2005) *Efeitos da dependência espacial em modelos de previsão de demanda por transporte*. Dissertação de Mestrado, Departamento de Transportes, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, 137 fl.
- Neter, J.; Wasserman, W.; Kutner, M.H. (1990) *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Design*. IRWIN, EUA.
- SPSS (2001) – *Statistical Package for the Social Sciences. Version 11.0 for Windows*. SPSS Inc, Illinois, USA.
- SAL (2004) *GeoDA – Geodata Analysis Software. Version 0.9.5-i (beta) for Windows*. Spatial Analysis Laboratory , Illinois, USA.

Universidade Federal do Ceará
Departamento de Engenharia de Transportes
Campus do Pici – Bloco 703 – CP 12.144 - CEP 60.455-760 – Fortaleza, CE

Fone/Fax: (0xx85) 4008-9488
E-mail: felipe@det.ufc.br
henriquensilva@gmail.com
carvalho@dam.brown.edu