

# MODELOS DE PREVISÃO DE DEMANDA POR TRANSPORTES EMPREGANDO ANÁLISE DE DEPENDÊNCIA ESPACIAL

**Simone Becker Lopes**

Universidade de São Paulo

Escola de Engenharia de São Carlos - Departamento de Transportes

**Antônio Néelson Rodrigues da Silva**

Universidade de São Paulo

Escola de Engenharia de São Carlos - Departamento de Transportes

## RESUMO

Este trabalho é uma contribuição para a investigação dos efeitos da dependência espacial em modelos de previsão de demanda por transportes. O objetivo principal é avaliar o desempenho de modelos tradicionais utilizados em análise de demanda por transportes e de modelos alternativos, nos quais são introduzidas variáveis de dependência espacial. A suposição básica deste estudo é que a introdução de indicadores de dependência espacial no processo de modelagem de demanda pode produzir estimativas mais acuradas do que aquelas obtidas com modelos tradicionais. O estudo de caso realizado para testar a referida hipótese foi baseado em dados de origem-destino (O-D) obtidos na cidade de Porto Alegre, RS, em dois períodos distintos. Os resultados obtidos mostram que o desempenho dos modelos em que foram introduzidas variáveis de dependência espacial são melhores que o dos modelos tradicionais. Adicionalmente, a distribuição espacial dos erros apresenta menos agrupamentos que a dos erros resultantes das estimativas dos modelos tradicionais.

## ABSTRACT

This work is a contribution to the investigation of the effects of spatial dependence on transportation demand models. Its main objective was to evaluate the performance of conventional mathematical models used in transportation demand analysis and of alternative models, in which spatial dependence variables were introduced. The basic assumption of this study was that the introduction of spatial dependence indicators in the demand modeling process could produce more accurate predictions than those obtained with traditional models. The case study carried out for testing the stated hypothesis was based on origin-destination (O-D) data of Porto Alegre, RS, in two different periods. The results obtained showed that the performance of models in which spatial dependence variables were introduced was better than that of traditional models. In addition, the spatial distribution of the errors was not as clustered as it was with the errors resulting from the estimations of traditional models.

## 1. INTRODUÇÃO

A maioria dos métodos e teorias estatísticas assume que as observações são independentes, condição, no entanto, raramente encontrada quando são analisados dados espaciais (Miller, 1999). Citando a Primeira Lei da Geografia, Tobler (1970) destaca que todos os objetos no espaço estão relacionados e que os resultados de análises de dados espaciais dependem da localização dos objetos que estão sendo considerados. Porém, estudos de demanda por transportes, indispensáveis ao planejamento de uma região ou cidade, têm sido gerados a partir de dados espacialmente dependentes, sem levar em consideração tal característica.

Grande parte do desenvolvimento analítico recentemente introduzido no campo do planejamento de transportes está associado aos Sistemas de Informações Geográficas (SIG), pois existem muitas questões analítico-espaciais que devem ser consideradas na análise de transportes, entre elas a definição de unidades de área modificáveis (zonas de tráfego), os problemas de contorno e amostragem espacial, a dependência e heterogeneidade espacial e as representações alternativas de ambientes geográficos.

O foco deste trabalho concentra-se na dependência espacial (também referida como autocorrelação espacial) para análise de dados de demanda por transportes, pois a técnica de análise de regressão, por exemplo, perde precisão sob os efeitos da dependência espacial, uma

vez que os parâmetros estimados não são eficientes e a significância dos testes é irreal na presença de autocorrelação espacial (Anselin e Griffith, 1988). Admite-se, também, que ignorar questões de análise espacial como, por exemplo, a dependência espacial, pode invalidar os resultados da análise da demanda por transportes, levar a uma previsão inadequada e conseqüentemente a um planejamento ineficiente.

Com o aumento da disponibilidade de bases digitais de dados geográficos e a capacidade de manuseio de informações geográficas, muitos estudos têm sido desenvolvidos sobre análise espacial nas últimas três décadas. No Brasil, apesar da crescente utilização das ferramentas de SIG, pouco tem sido pesquisado sobre o assunto para a modelagem de transportes (como Krempi *et al.*, 2002; Teixeira, 2003; Queiroz, 2003; Krempi, 2004; Henrique, 2004; Lopes e Silva, 2004 e Lopes, 2005).

O objetivo principal deste trabalho é avaliar, mediante o uso de ferramentas de planejamento de transportes e de análise espacial em ambiente SIG, os efeitos da utilização de indicadores de dependência espacial em modelos de previsão de demanda por transportes, ou seja, verificar se a proposição de modelos alternativos, com variáveis espaciais, produz melhores resultados que os obtidos por modelos estatísticos tradicionais.

## 2. FERRAMENTAS DE ANÁLISE EXPLORATÓRIA DE DADOS ESPACIAIS

As técnicas de análise exploratória aplicadas a dados espaciais são essenciais ao desenvolvimento da modelagem em estatística espacial. Consistem em adaptações de ferramentas usuais, sendo que Anselin (1996) refere-se a elas como ESDA (*Exploratory Spatial Data Analysis*), uma vez que podem ser consideradas uma extensão da EDA (*Exploratory Data Analysis*). As técnicas ESDA são apresentadas na forma de índices, que medem a associação espacial (Índice de Moran), gráfico de espalhamento (*Moran Scatterplot*) e mapas (*Box map*). Sua função é auxiliar na identificação de agrupamentos de objetos, cujos atributos possuem valores altos e baixos, áreas de transição e casos atípicos. Em toda técnica ESDA são encontrados três elementos básicos:

- Matriz de proximidade espacial ( $W$ ): matriz quadrada que estima a variabilidade espacial de dados de área, em que cada elemento  $w_{ij}$  representa uma medida de proximidade entre  $A_i$  e  $A_j$ , sendo  $A_i$  e  $A_j$  as zonas que estão sendo analisadas;
- Vetor dos desvios ( $Z$ ): calculado a partir da média geral ( $\mu$ ), cada elemento é calculado subtraindo-se o valor de  $\mu$  do valor do atributo ( $y_i$ ) para cada região;
- Vetor de médias ponderadas ( $W_z$ ): também chamado de média dos valores dos vizinhos ou média móvel espacial, é uma medida útil para o cálculo da variação da tendência espacial.

Segundo Câmara *et al.* (2002), um aspecto fundamental na análise exploratória espacial é a caracterização da dependência espacial, que mostra como os valores estão correlacionados no espaço. As funções utilizadas para estimar quanto o valor observado de um atributo de uma região é dependente dos valores dessa mesma variável nas localizações vizinhas são a *autocorrelação espacial* e o *variograma*. Uma das formas de medir a autocorrelação espacial é através do índice  $I$  de Moran, cujo valor varia de  $-1$  a  $+1$  e indica quanto cada área analisada é semelhante à sua vizinhança imediata (Equação 1):

$$I = \frac{Z^t \times W_z}{Z^t \times Z} \quad (1)$$

onde

$I$  = índice de autocorrelação espacial;

$Z^t$  = vetor transposto dos desvios.

Uma maneira de visualizar a dependência espacial é através do diagrama de espalhamento de Moran. Construído com base nos valores normalizados ( $Z$ ) e comparados com a média dos vizinhos ( $W_z$ ), em um gráfico bidimensional dividido em quatro quadrantes, permite analisar o comportamento da variabilidade espacial. O índice  $I$  de Moran é equivalente ao coeficiente de regressão linear, que indica a inclinação da reta de regressão ( $\alpha$ ) de  $W_z$  em  $Z$ , e os quadrantes podem ser interpretados como:

- $Q1$  (valores positivos e médias positivas) e  $Q2$  (valores negativos e médias negativas): indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes.
- $Q3$  (valores positivos e médias negativas) e  $Q4$  (valores negativos e médias positivas): indicam pontos de associação espacial negativa, no sentido que uma localização possui vizinhos com valores distintos.

O diagrama de espalhamento de Moran também pode ser apresentado na forma de um mapa temático bidimensional (*Box Map*), no qual cada polígono é apresentado indicando-se seu quadrante no diagrama de espalhamento. Enquanto os indicadores globais, como o índice  $I$  de Moran, fornecem um único valor como medida de associação espacial para todo o conjunto de dados, os indicadores locais produzem um valor específico para cada área, permitindo, assim, a identificação de agrupamentos de áreas com valores de atributos semelhantes (*clusters*), valores anômalos (*outliers*) e de mais de um regime espacial (INPE, 2004). Anselin (1996) refere-se a estes indicadores como estatística LISA (*Local Indicators of Spatial Association*), os quais devem satisfazer dois requisitos básicos:

- Permitir a identificação de padrões significativos de associação espacial;
- A soma dos indicadores LISA, para todas as observações, deve ser proporcional ao índice global de associação espacial.

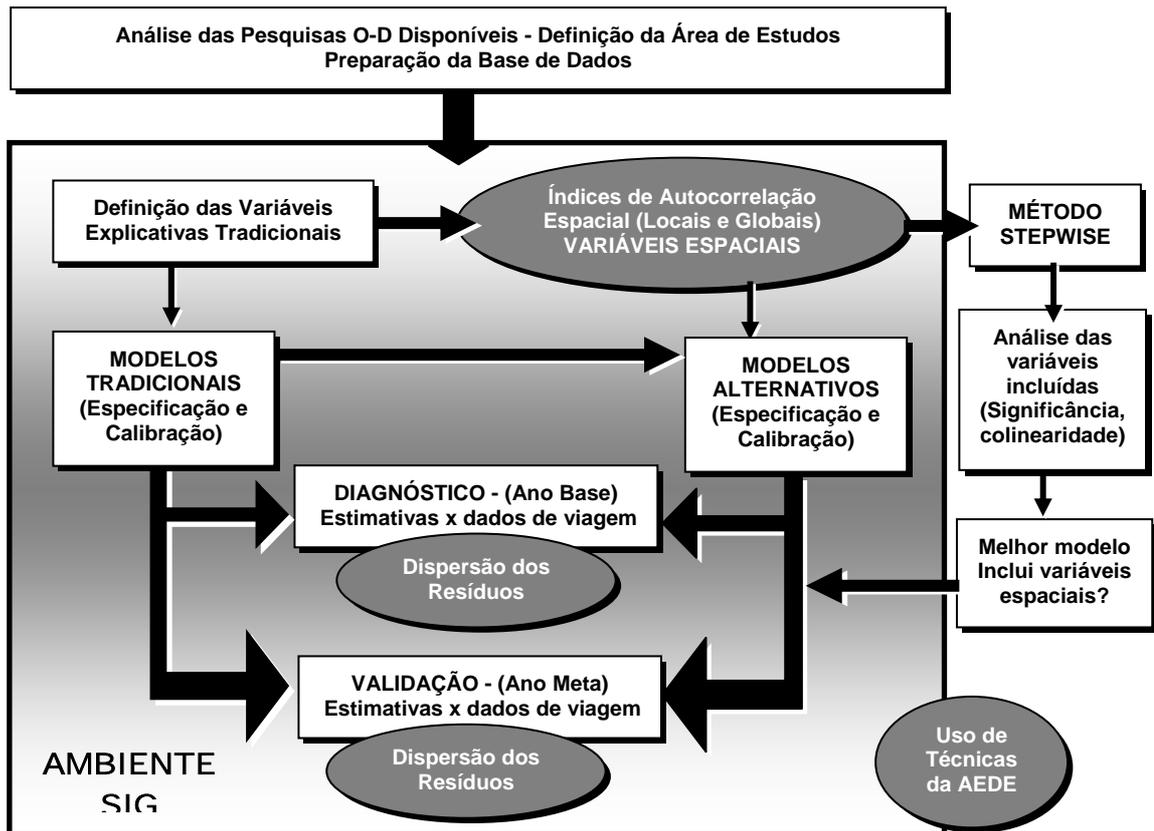
A significância estatística do uso do índice de Moran local é computada de forma similar ao caso do índice global. Depois de calculado o índice para cada área, permuta-se, aleatoriamente, o valor das demais áreas, até ser obtida uma pseudo-distribuição para a qual se possa computar os parâmetros de significância. É útil gerar mapas (*LISA Map* e *Moran Map*) indicando as regiões que apresentam correlação local significativamente diferente do resto dos dados, pois são áreas com dinâmica espacial própria, ou seja, “bolsões” de não-estacionaridade, e merecem análise detalhada. Autocorrelação com significância superior a 95 % indica, a princípio, áreas com grande semelhança em relação a seus vizinhos.

### 3. MÉTODO

As etapas que constituem o método aplicado para atingir os objetivos propostos, assim como a interrelação entre elas, são apresentadas na Figura 1. Estão destacadas as etapas que envolvem o uso de ferramentas de Análise Exploratória de Dados Espaciais (AEDE), salientando-se, também, que praticamente todo o processo ocorre em ambiente SIG.

Inicialmente, deve-se ressaltar que este estudo tem por base dados de pesquisas de origem e destino (O-D) realizadas na cidade de Porto Alegre-RS em dois períodos distintos:

- “Ano base” – dados utilizados para calibração de modelos de demanda por transportes e também para diagnóstico dos melhores modelos, ou seja, daqueles que são usados para estimativas de viagens futuras, tanto por métodos tradicionais (**modelos tradicionais**) como através de métodos que introduzem variáveis que medem o grau de dependência espacial (**modelos alternativos**);
- “Ano meta” – dados da pesquisa do ano de projeção das viagens, que servem de base para a avaliação da qualidade das estimativas futuras de viagens obtidas através da aplicação tanto dos **modelos tradicionais** como dos **modelos alternativos**.



**Figura 1:** Esquema do método aplicado, com destaque para as etapas de uso da AEDE

Conforme apresentado na Figura 1, as técnicas da AEDE são aplicadas não somente para obtenção de indicadores de dependência espacial, necessários para a especificação dos modelos alternativos, como também no diagnóstico dos modelos (tradicionais e alternativos), auxiliando na seleção daqueles que serão aplicados para estimativas futuras (validação), e, finalmente, na análise comparativa dos resultados.

Além das ferramentas disponíveis no SIG-T, a inclusão de variáveis espaciais foi analisada através da utilização do método de *Regressão Stepwise Forward*, que verifica se a adição de uma nova variável ao modelo provoca um aumento significativo no coeficiente de determinação ajustado ( $R^2_{aj}$ ). Deve-se destacar, no entanto, que o método não prescinde da avaliação do modelo resultante pelo analista, uma vez que, em alguns casos, as ferramentas utilizadas podem não conseguir interpretar o problema de colinearidade.

#### 4. APLICAÇÃO DO MÉTODO E ANÁLISE DOS RESULTADOS

O estudo de caso realizado para testar os diferentes modelos baseou-se em dados de pesquisa de origem-destino (O-D) obtidos através de entrevista domiciliar (EDOM) realizada na cidade de Porto Alegre nos anos de 1974 e 1986. Duas outras pesquisas O-D foram realizadas na cidade ou na região metropolitana depois de 1986, mas não puderam ser utilizadas pelas razões apresentadas por Lopes e Silva (2004) e por Lopes (2005). Desta forma, os dados de 1974 constituíram os dados do ano base e os dados de 1986 os dados do ano meta.

Porto Alegre, que é a principal cidade e capital do estado do Rio Grande do Sul, apresentava uma população em torno de um milhão de habitantes em 1974, 20 % maior em 1986 e pouco acima de 1,36 milhões de habitantes em 2000, ano do último recenseamento (IBGE, 2000). As informações, agrupadas em Zonas de Tráfego (ZTs), referem-se não somente às características das viagens, mas também a características socioeconômicas importantes dos entrevistados: distribuição espacial da população, renda, posse de veículos particulares, uso do solo, distribuição das atividades econômicas etc.

Os dados das 96 ZTs de 1974 precisaram ser inteiramente digitalizados e, posteriormente, as duas bases de dados, incluindo agora 75 das 95 ZTs de 1986, foram transferidas para um Sistema de Informações Geográficas para Transportes (SIG-T) e também para um SIG que contém as ferramentas de Análise Exploratória de Dados Espaciais.

Deve-se salientar que o foco do estudo se restringe apenas à fase de estimativa de *Viagens Produzidas de Base Domiciliar* (VPBD), que é apenas uma parte da primeira etapa do Modelo de Quatro Etapas. Além disso, as viagens estão agregadas em todos os modos e motivos por falta de detalhamento nos relatórios da pesquisa do ano base. Dessa forma, o método aplicado não pretende esgotar o assunto, apenas dar início a uma investigação que se insere no extenso e ainda pouco explorado tema de Análise Espacial, apresentando um exemplo de aplicação e verificando sua relevância para o Planejamento de Transportes.

Cabe ressaltar que o estudo ocorreu em **três fases**. Essas fases foram acontecendo como uma forma de aperfeiçoamento da pesquisa, sempre em busca de outras variáveis explicativas (tanto tradicionais como alternativas), com a finalidade de encontrar o melhor modelo tradicional possível e o melhor modelo alternativo possível, utilizando os dados disponíveis da pesquisa O-D do ano base.

A **Fase 1** foi um estudo preliminar (Lopes e Silva, 2004), em que apenas algumas etapas foram desenvolvidas, com o intuito de verificar a validade da hipótese básica, ou seja, que a introdução de indicadores de dependência espacial no processo de modelagem de demanda por transportes poderia produzir resultados mais acurados que os obtidos com modelos tradicionais. Também serviu para a definição de diretrizes para a segunda fase.

Na **Fase 2** foi introduzida a análise da estatística LISA (indicadores locais de autocorrelação espacial), não abordada na fase preliminar. Foi feita uma análise mais detalhada dos modelos, com análise de significância das variáveis incluídas através de ferramentas de SIG-T e também de técnicas estatísticas tradicionais (em particular o método de regressão *stepwise*).

Na **Fase 3** foram geradas e incluídas nos modelos outras variáveis, sob a forma de “taxas”. Foram adotados critérios mais rigorosos na especificação e diagnóstico dos modelos e também na análise das variáveis explicativas.

Para a definição das variáveis independentes que entrariam nos modelos foram selecionadas, na **Fase 2**, sete variáveis, entre as disponíveis na base de dados de 1974, consideradas importantes para a modelagem de *Viagens Produzidas de Base Domiciliar* (VPBD): *AREA* (km<sup>2</sup>); *DENS* (habitantes/km<sup>2</sup>); *DOMIC* (domicílios/ZT); *POP\_T* (população total); *POP\_A* (população ativa); *RENDA* (média da renda mensal dos domicílios) e *FROTA* (número de veículos particulares). Foi realizada análise de correlação linear dessas variáveis com **VPBD**.

Inicialmente, foram analisados *digramas de dispersão* e verificou-se a existência de relação linear, não havendo necessidade de transformação de variáveis. Uma *matriz de correlação* foi gerada através de ferramenta de estatística do **SIG-T**, utilizando-se a base de dados da EDOM 1974 (**ano base**). Foram analisados os valores dos *coeficientes de correlação R* entre as variáveis socioeconômicas (potenciais variáveis explicativas) e a variável dependente **VPBD**. O maior índice foi verificado para a variável *POP\_T* (R = 0,91), seguida pelas variáveis *DOMIC*, *POP\_A* e *FROTA* (R = 0,90; R = 0,87 e R = 0,84, respectivamente). Como as três primeiras também apresentaram correlação alta entre si (R > 0,95), sua inclusão no modelo teve que ser feita de maneira criteriosa, para evitar possíveis problemas de multicolinearidade.

Para a especificação dos modelos tradicionais foram testadas 21 combinações das 7 variáveis, usando-se os dados de 90 das 96 ZTs da pesquisa O-D de 1974, através das ferramentas de planejamento do **SIG-T**. Seis ZTs foram excluídas por representarem zonas atípicas (Lopes 2005). O diagnóstico dos modelos para a escolha do melhor modelo tradicional foi baseado na análise das seguintes medidas de desempenho: *coeficiente de determinação (R<sup>2</sup>)*; *coeficiente de determinação ajustado (R<sup>2</sup><sub>aj</sub>)*, indicado quando há modelos com diferentes números de variáveis incluídas; *estatística t (T stat)* para cada variável; e, finalmente, os *coeficientes das variáveis*, com verificação de coerência entre atributo e variável dependente, pois um coeficiente com sinal invertido pode indicar presença de multicolinearidade.

Após a análise dessas medidas de desempenho, fornecidas nos relatórios gerados para cada modelo calibrado, obteve-se um modelo tradicional, denominado modelo **T**, que incluiu apenas duas das sete variáveis analisadas (*POP\_T* e *FROTA*) e que apresentou valores elevados para **R<sup>2</sup>** (0,916) e **R<sup>2</sup><sub>aj</sub>** (0,914). Na última etapa do diagnóstico foram avaliados os resultados das estimativas do modelo **T** para o **ano base**, através dos valores de **ERM (Erros Relativos Médios)** e **I\_Mor\_Res (I de Moran dos Resíduos)**, usando-se como referência os dados de viagem de 1974. Para a validação, o modelo **T** foi aplicado para estimativas futuras (**ano meta**) e as mesmas medidas foram avaliadas com base nos dados de viagem de 1986.

O valor para o **ERM** das estimativas futuras apresentou-se relativamente alto (57 %), enquanto que o resultado das estimativas para o **ano base** teve **ERM** de 12 %, problema também constatado por Lopes e Silva (2004). Quanto à análise da dispersão dos resíduos, o valor do **I\_Mor\_Res** (0,39) nas estimativas para o **ano base** indica presença de autocorrelação espacial, ou seja, tendência de concentração de resíduos positivos ou negativos numa região do mapa, o que não foi verificado, no entanto, na estimativa futura, apresentando valor próximo de zero para o **I\_Mor\_Res** (0,09). A grande amplitude dos valores das variáveis incluídas pode ser a causa das estimativas para o **ano meta** apresentarem maiores

erros do que as estimativas para o **ano base** (Lopes, 2005), pois, nesses casos, os coeficientes ajustados podem ser influenciados pela alta variância, ficando sub ou superestimados.

Esse fato é uma explicação possível para o modelo ajustado não ter apresentado uma boa adequação aos dados de 1986. Como tentativa de correção do problema foi feita, então, a padronização das variáveis, sendo calibrado um novo modelo, similar ao modelo **T**, porém com as variáveis padronizadas ( $POP_{T_p}$  e  $FROTA_p$ ). As características e resultados obtidos com o modelo **T<sub>p</sub>** são apresentados na Tabela 1, destacando-se que os erros nas estimativas para o **ano meta** tornaram-se significativamente menores que os do modelo **T**.

O resultado referente à estimativa para o **ano base** permaneceu igual ao apresentado pelo modelo **T**, assim como os valores de  $R^2$  e  $R^2_{aj}$ . Porém, como o efeito da padronização foi significativo quando da análise das estimativas futuras, admite-se que o modelo **T<sub>p</sub>** é superior ao modelo **T**, apesar de ainda apresentar um índice muito alto de autocorrelação espacial dos resíduos das estimativas para o **ano base** (0,39).

Na **Fase 3** do estudo foram definidas novas variáveis (**taxas**) para serem incluídas nas análises:  $X21$  (DOMIC/AREA),  $X32$  ( $POP_{T/DOMIC}$ );  $X51$  ( $FROTA/AREA$ );  $X53$  ( $FROTA/POP_{T}$ ). Procurou-se testar a hipótese apresentada por Teixeira (2003) de que as variáveis em forma de “taxas” podem amenizar o *Problema de Unidades de Áreas Modificáveis* (PUAM ou MAUP, na sigla em inglês), que é um problema inerente aos dados de áreas, onde as estimativas obtidas dentro de um sistema de unidade de área dependem das diferentes maneiras que essas áreas podem ser agregadas, ou seja, podem ser obtidos diferentes resultados apenas mudando-se as fronteiras de determinadas zonas.

No entanto, as variáveis (taxas) não apresentaram bons resultados para os modelos tradicionais. Já de início, verificaram-se menores *índices de correlação* com a variável dependente do que os verificados com as variáveis de contagens agregadas. Além disso, apresentaram altos índices de correlação entre si, indicando que a inclusão de mais de uma delas no modelo possivelmente geraria problemas de multicolinearidade. Essa suposição foi confirmada na etapa de especificação dos modelos, através da análise dos valores da *estatística t* e dos *coeficientes*, verificando-se que nenhuma delas se mostrou significativa para os modelos tradicionais. Concluiu-se que o melhor modelo tradicional, considerando-se as variáveis disponíveis na pesquisa de 1974, é o modelo **T<sub>p</sub>**, obtido na **Fase 2**.

O próximo passo foi a especificação de modelos alternativos correspondentes, através da introdução das **variáveis espaciais (Globais e Locais)**, definidas pela análise espacial das variáveis socioeconômicas e verificação das que apresentam autocorrelação espacial significativa. Na **Fase 3**, a definição das variáveis espaciais incluiu, ainda, a análise de correlação da distribuição espacial das variáveis com a distribuição espacial de VPBD.

Os maiores índices de autocorrelação espacial entre as variáveis analisadas na **Fase 2**, para os dados de 1974, foram para *REND*A e *DENSI*, que apresentaram índices *I de Moran* de 0,5881 e 0,7330, respectivamente. Os valores correspondentes para os dados de 1986 foram 0,4404 e 0,5681. Nas análises da **Fase 3** destacaram-se as variáveis  $X21$  e  $X51$ , com índices *I de Moran* de 0,7860 e 0,7543, para os dados de 1974, e 0,6428 e 0,6279 para os dados de 1986, respectivamente (Tabela 2). Esses valores indicam um grau razoável de autocorrelação espacial, o que foi confirmado pela análise visual de *Box Maps* (Lopes e Silva, 2004).

Verifica-se, na análise da Tabela 2, que os maiores índices de autocorrelação espacial são observados para as variáveis em forma de taxas, particularmente nas que utilizam unidades de área (*DENSI*, *X21* e *X51*). É uma evidência de que a autocorrelação espacial das contagens agregadas de população, domicílio e frota não foram devidamente medidas, pois estavam sob o efeito da *MAUP*. Confirma, também, a hipótese de que a transformação de variáveis em taxas por unidade de área diminui as distorções do zoneamento e os efeitos da *MAUP*, que não tinham ficado evidentes inicialmente.

Para considerar a dependência espacial nos modelos foram definidas *variáveis espaciais globais* e *variáveis espaciais locais*. As globais foram obtidas através da especificação de variáveis binárias (*dummy*) associadas aos quadrantes do *diagrama de Moran* (indicador global). Para as análises da **Fase 2**, foram definidas seis variáveis que representam a autocorrelação espacial global: três *variáveis espaciais globais* para *DENSI* (*D\_Q1*, *D\_Q2* e *D\_Q3*) e três para *REND*A (*R\_Q1*, *R\_Q2* e *R\_Q3*). Para a definição das *variáveis espaciais locais*, foram considerados os indicadores locais de autocorrelação espacial (*LISA*) para as mesmas variáveis (*DENSI* e *REND*A). Dessa forma, duas variáveis espaciais (*LISA\_D* e *LISA\_R*) representam os índices locais de autocorrelação espacial (*I* de Moran Local) de cada *ZT* das variáveis *densidade populacional* e *renda média mensal*, respectivamente.

Na **Fase 3**, as análises comparativas da distribuição espacial das variáveis socioeconômicas com a variável dependente **VPBD** revelaram a existência de outras variáveis espaciais, que poderiam ser potenciais variáveis explicativas para o fenômeno em estudo. Foram verificadas as *ZTs* com quadrantes coincidentes das variáveis socioeconômicas com os quadrantes da variável dependente na análise comparativa dos *Box Maps* e *Moran Maps*. Selecionou-se, então, para especificação dos modelos alternativos da **Fase 3**, 12 *variáveis espaciais globais* e 4 *variáveis espaciais locais*, derivadas da análise espacial das variáveis *DENSI*, *X21*, *X51* e *X53*, que foram as que mais apresentaram correlação espacial com a variável dependente.

A especificação dos modelos alternativos, para as duas fases, foi a partir do modelo tradicional selecionado, **T<sub>p</sub>**, com a introdução das *variáveis espaciais* de cada fase, em três diferentes métodos: introdução somente de variáveis espaciais globais (modelo “AG”), através da introdução de variáveis espaciais locais (modelo “AL”) e também através da combinação de variáveis espaciais globais e locais (modelo “AGL”). Foram considerados os mesmos critérios de análise de significância das variáveis, ou seja, análise dos valores de **t stat** e da coerência dos coeficientes, que indicaram apenas dois modelos alternativos como válidos na **Fase 2**: o modelo **AL<sub>p</sub>** e o modelo **AGL<sub>p</sub>**. Na **Fase 3** foram obtidos três modelos alternativos, um de cada tipo, que foram chamados de **F3AG<sub>p</sub>**, **F3AL<sub>p</sub>** e **F3AGL<sub>p</sub>**.

Os modelos alternativos, da mesma forma que o tradicional **T<sub>p</sub>**, foram utilizados para estimativas para o ano base. Os resultados são evidências de que a introdução das variáveis espaciais melhora o desempenho dos modelos, pois a maioria das medidas analisadas melhorou com os modelos alternativos quando comparadas com os resultados do modelo **T<sub>p</sub>**. Foi escolhido o modelo **AGL<sub>p</sub>** como melhor modelo alternativo, na **Fase 2**, pois apesar de não apresentar alteração no valor de *ERM* em relação ao modelo tradicional, apresentou os melhores resultados nas outras três medidas analisadas, destacando-se o valor do **I de Mor Res**, que baixou de 0,39 para 0,17, representando 56 % de melhora relativa (Tabela 1).

Na **Fase 3** o modelo escolhido como o melhor modelo alternativo foi, mais uma vez, o do tipo “AGL”. Os resultados apresentados com o modelo alternativo **F3AGL<sub>p</sub>** superaram, inclusive, os obtidos com o modelo alternativo da **Fase 2**, conforme pode ser verificado na Tabela 1. Destacou-se, mais uma vez, o valor do **I de Mor Res**, que baixou para 0,10, representando melhora relativa de 74 % em relação ao modelo tradicional (0,39). Observou-se, também, melhora relativa de 17 % no valor do ERM.

Anteriormente à etapa de validação dos modelos, como uma forma complementar ao procedimento aplicado e verificação de que as variáveis incluídas eram realmente as mais significativas, uma vez que foram muitas as analisadas, especificou-se um modelo através do método de *Regressão Stepwise Forward*. Foram incluídas, para análise através do *software* SPSS, as 51 variáveis estudadas (sendo 11 tradicionais e 40 espaciais). Como resultado obteve-se o mesmo modelo alternativo selecionado como o melhor na **Fase 3** (**F3AGL<sub>p</sub>**), confirmando a validade dos procedimentos adotados com o uso das ferramentas de SIG-T.

Finalmente, os modelos **AGL<sub>p</sub>** e **F3AGL<sub>p</sub>** são aplicados para estimativas futuras. As características desses modelos alternativos, bem como as do modelo tradicional **T<sub>p</sub>**, e os resultados obtidos na etapa de validação podem ser observados na Tabela 1. Destaca-se o modelo alternativo da **Fase 2** (**AGL<sub>p</sub>**), com o valor mais baixo para o **I\_Mor Res** (0,04), numa melhora de 55,6 % em relação ao resultado obtido com o modelo tradicional **T<sub>p</sub>** (0,09). Tem-se um melhor entendimento desse resultado na análise do *Moran Map* (Figura 2).

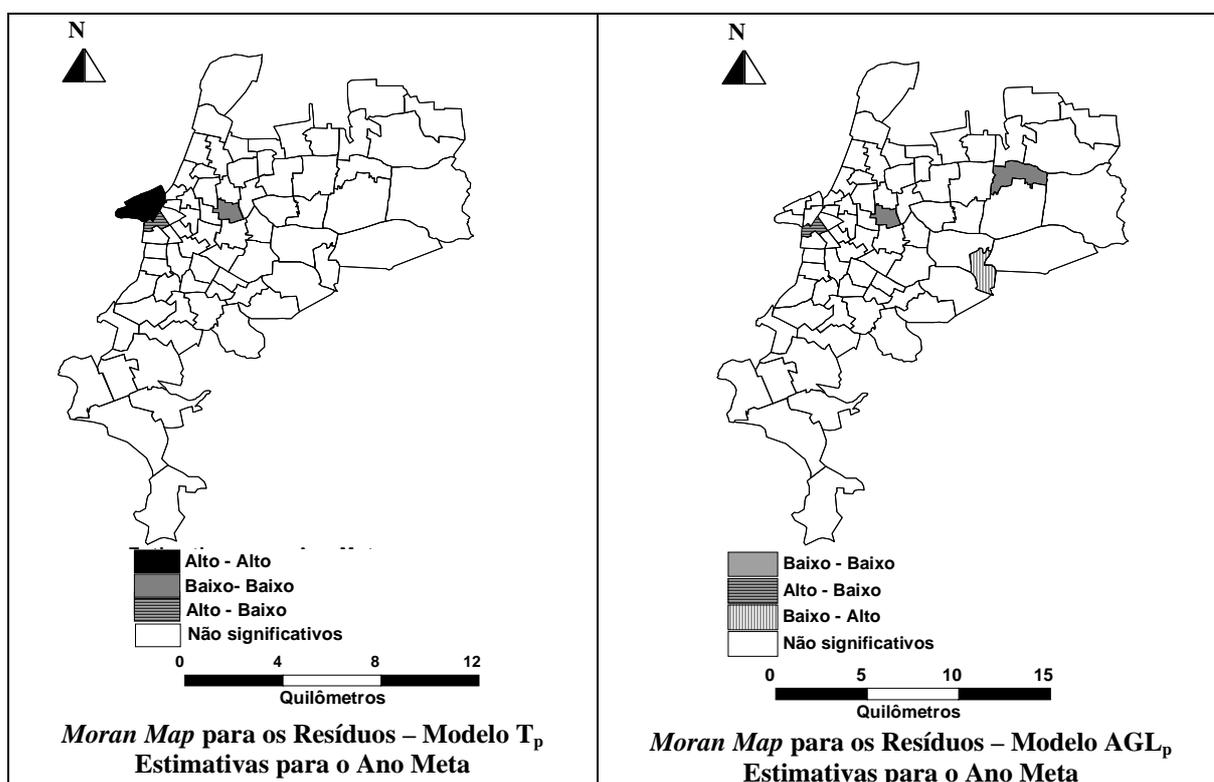
**Tabela 1:** Características e resultados obtidos com os modelos T<sub>p</sub> e AGL<sub>p</sub>

		Modelos para (VPBD) analisados						
		<i>T<sub>p</sub></i>		<i>AGL<sub>p</sub></i>		<i>F3AGL<sub>p</sub></i>		
Variáveis explicativas	Tradicionais	Constante	12195,4	11587,7		12428,00		
		<i>POP_Tp</i>	3911,26	3683,05		4053,69		
		<i>FROTA<sub>p</sub></i>	2576,07	2479,25		2108,89		
	Espaciais	<i>D_Q1</i>		1709,20		1003,06		
		<i>X51_Q2</i>				-1056,87		
		<i>LISA_D</i>		-1079,25				
		<i>LISA_R</i>		-356,59				
		<i>LISA_X21</i>				-1264,40		
	Modelo		<b>R<sup>2</sup></b>	0,916	0,949	3,6%	0,955	4,3%
			<b>R<sup>2</sup><sub>aj</sub></b>	0,914	0,946	3,5%	0,952	4,2%
Diagnóstico	Estimativa (ano meta)	<i>ERM</i>	12 %	12 %	0%	10 %	16,7%	
		<i>I_Mor_Res</i>	0,39	0,17	56,4%	0,10	74,4%	
Validação	Estimativa (ano meta)	<i>ERM</i>	13 %	13 %	0	14 %	-7,7%	
		<i>I_Mor_Res</i>	0,09	0,04	55,6%	0,11	-22,2 %	

Pode-se verificar, também, nos resultados da Tabela 1, que, para estimativas futuras, ambos os modelos (T<sub>p</sub> e AGL<sub>p</sub>) apresentaram valores iguais para os **ERM** (13 %), sendo um pouco acima do valor obtido na etapa de diagnóstico. Observa-se, no entanto que o modelo que apresentou melhores resultados na etapa de diagnóstico (**F3AGL<sub>p</sub>**) foi o que apresentou os piores resultados na etapa de validação, apresentando em relação ao modelo tradicional variação de -8% para o ERM e -22% para o **I\_Mor Res**.

**Tabela 2:** Índices de autocorrelação espacial (*I* de Moran Global) das variáveis socioeconômicas e da variável dependente VPBD para os dados do **ano base** (1974) e do **ano meta** (1986)

Variáveis		Descrição da Variável	I de Moran Global			
			1974	1986		
<b>Dependente</b>	<b>VPBD</b>	Viagens produzidas de base domiciliar	0,3467	0,2849		
	<b>Totais de contagem</b>	<b>DOMIC</b>	Número total de domicílios	0,1881	0,1216	
<b>Independentes</b>		<b>POP_T</b>	Número total de habitantes	0,2352	0,2379	
		<b>FROTA</b>	Número total de veículos particulares	0,2300	0,1990	
	<b>Fase 2</b>	<b>RENDA</b>	Renda média mensal por domicílio	0,5881	0,4404	
		<b>DENSI</b>	Densidade (habitantes por km <sup>2</sup> )	0,7330	0,5681	
	<b>Taxas derivadas</b>	<b>X21</b>	Número de domicílios por Km <sup>2</sup>	0,7860	0,6428	
		<b>X32</b>	Habitantes por domicílio	0,3784	0,5349	
		<b>Fase 3</b>	<b>X51</b>	Veículos particulares por km <sup>2</sup>	0,7543	0,6279
			<b>X52</b>	Veículos particulares por domicílio	0,4469	0,2584
<b>X53</b>			Veículos particulares por habitante	0,2997	0,4321	



**Figura 2:** Moran Map para os resíduos das estimativas para o **ano meta**, com o modelo tradicional  $T_p$  e com o modelo alternativo  $AGL_p$

Observa-se, na análise da dispersão dos resíduos através do *Moran Map* (Figura 2), um agrupamento com significância acima de 95 %, constituído por 4 ZTs no centro da cidade, que apresentaram valores altos de resíduos (Alto–Alto) nas estimativas futuras com o modelo tradicional  $T_p$ . Esse agrupamento, que não é verificado para os resíduos com o modelo alternativo  $AGL_p$ , representa ZTs com valores baixos de viagens em 1986 (**ano meta**) e que foram superestimados com o modelo tradicional. As outras duas ZTs em agrupamentos

significativos, verificadas com o modelo  $T_p$ , são as mesmas apresentadas para o modelo  $AGL_p$  (sendo uma Baixo-Baixo e uma Alto-Baixo).

## 5. CONCLUSÃO

No estudo apresentado neste artigo, para a análise comparativa de modelos tradicionais com modelos que introduzem variáveis espaciais, foi considerada a variável dependente *Viagens Produzidas de Base Domiciliar (VPBD)*, agregada em todos os modos e todos os motivos, pois a pesquisa O-D de 1974 (ano-base) não apresentava resultados de viagens desagregados. Conseqüentemente, este trabalho é antes de tudo um exemplo de aplicação do método, cujos resultados evidenciam a importância da investigação de outras etapas da projeção de demanda por transportes, assim como encorajam a análise de formas mais desagregadas de análise da variável dependente (por modo ou por motivo de viagem).

Na análise dos modelos tradicionais, destaca-se particularmente o efeito da padronização de variáveis para a obtenção de melhores modelos tradicionais. A introdução de variáveis em forma de taxas derivadas não teve efeito significativo, mas confirmou a hipótese de que poderia amenizar o problema inerente aos dados de áreas (MAUP), o que foi verificado através da análise espacial, onde o efeito da autocorrelação espacial se mostrou mais evidente justamente nas taxas por unidade de área. Foram essas variáveis derivadas que conseguiram revelar o efeito da dependência espacial, possivelmente camuflado por um zoneamento de áreas de dimensões muito diferenciadas, que afetava as variáveis de contagens agregadas.

A hipótese básica do estudo foi confirmada, pois a introdução de variáveis espaciais na análise de modelos de demanda por transporte apresentou resultados positivos, em relação aos modelos tradicionais, tanto na etapa de diagnóstico, quando aplicados para os dados do ano base, como na etapa de validação, com estimativas para o ano meta. No entanto, requer, da mesma forma que os modelos tradicionais, rigor na análise de significância das variáveis. Dentre os métodos analisados, verificou-se que a combinação dos dois tipos de variáveis espaciais (globais e locais) foi mais eficiente do que a consideração dessas variáveis isoladamente. Provavelmente porque os modelos resultantes conseguem contemplar os efeitos da dependência espacial de todas variáveis, evitando, porém, o problema de multicolinearidade entre elas, uma vez que são introduzidas através de diferentes indicadores.

O fato de o modelo alternativo mais ajustado para os dados do ano base não ter sido o mesmo que apresentou os melhores resultados para as estimativas futuras conduz à conclusão complementar de que a dinâmica apresentada pelo desenvolvimento urbano, como é o caso de Porto Alegre, acarreta alterações nas relações entre as diferentes variáveis com o fenômeno estudado, modificando, inclusive, os padrões espaciais. Considera-se, portanto, que a análise dessa dinâmica e o estudo de formas de introduzi-la nos modelos de demanda por transportes podem produzir resultados ainda melhores.

Os resultados obtidos com o algoritmo *stepwise* comprovaram a importância da introdução de variáveis espaciais e, também, a validade do método empregado para especificação dos modelos através de ferramentas do SIG-T, uma vez que o modelo obtido foi o mesmo modelo alternativo considerado como o melhor modelo na etapa de diagnóstico.

Deve-se destacar, finalmente, que, apesar da contribuição importante de ferramentas de estatística neste estudo, a base das análises foi constituída por ferramentas de planejamento de

transportes e de análise e estatística espacial em ambiente SIG, o que salienta a importância dessa plataforma na análise de demanda por transportes. Além de facilitar a manipulação de grande quantidade de dados e a identificação e correção de erros, permite a incorporação de características espaciais através de ferramentas de análise estatística espacial. Possibilita, também, a integração de gráficos a mapas, permitindo a análise visual dos resultados.

#### AGRADECIMENTOS

Os autores agradecem à CAPES, CNPq e FAPESP pelos auxílios concedidos para a realização de pesquisas que deram origem a este trabalho.

#### REFERÊNCIAS BIBLIOGRÁFICAS

- Anselin, L. (1995) Local indicators of spatial association - LISA. *Geographical Analysis*, 27: 91-114
- Anselin, L. (1996) The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In M. Fischer, H. Scholten and D. Unwin (eds.), *Spatial Analytical Perspectives on GIS*. Taylor & Francis, London.
- Anselin, L. e D. A. Griffith (1988) Do spatial effects really matter in regression analysis? *Papers of the Regional Science Association* 65, 11-34.
- Câmara, G.; M. S. Carvalho; O. G. Cruz, V. Correa, (2002). Análise Espacial de Áreas, In: 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 – Disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/> Acesso em: 23 de março de 2004
- Henrique, C. S. (2004). Diagnóstico Espacial da Mobilidade e da Acessibilidade dos Usuários do Sistema Integrado de Transporte de Fortaleza. Dissertação de Mestrado, Universidade Federal do Ceará, Fortaleza.
- IBGE (2000) Censo Demográfico 2000 - Resultados do Universo - Porto Alegre. Acesso em 10 de Maio de 2004 (<http://www.ibge.gov.br/english/estatistica/populacao/censo2000/universo.php?tipo=31&uf=43>).
- INPE (2004) SPRING - Sistema de Processamento de Informações Georreferenciadas - Instituto Nacional de Pesquisas Espaciais – Divisão de Processamento de Imagens, São José dos Campos, São Paulo. Disponível em: <http://www.dpi.inpe.br/spring/portugues/download.php>. Acesso em: 20 de março de 2004
- Krempi, A. P. (2004) Explorando Recursos de Estatística Espacial para Análise da Acessibilidade de Cidade de Bauru. Dissertação de Mestrado, Universidade de São Paulo, São Carlos, SP, Brasil.
- Krempi, A. P., N. C. M. Brondino, A. N. R. Silva (2002) Evaluating transportation accessibility with spatial statistics tools in a GIS environment. In: *Proceedings of International Conference on Design and Decision Support Systems in Architecture and Urban Planning*, 6, Ellecom – Holanda, 2002. *Proceedings...* Eindhoven, Eindhoven University of Technology, p. 132-144.
- Lopes, S. B. (2005) Efeitos da Dependência Espacial em Modelos de Previsão de Demanda por Transporte. Dissertação de Mestrado, Universidade de São Paulo, São Carlos, SP, Brasil.
- Lopes, S. B. e A. N. R. Silva (2004) An Assessment Study of the Spatial Dependence in Transportation Demand Models. In: *Anais do Congresso Panamericano de Engenharia de Trânsito e Transportes*, 13, Albany, NY, EUA, 26 a 29 de Setembro.
- Miller, H. J. (1999) Potential contribution of spatial analysis to Geographic Information Systems for Transportation (GIS-T) *Geographical Analysis*, 31, 373-399.
- Queiroz, M. P. (2003) Análise Espacial dos Acidentes de Trânsito do Município de Fortaleza -Dissertação de Mestrado – UFC – Fortaleza – CE – 124p.
- Teixeira, G. L. (2003) Uso de Dados Censitários para Identificação de Zonas Homogêneas para Planejamento de Transportes Utilizando Estatística Espacial – Dissertação de Mestrado – Publicação T.DM-010A/03 – Departamento de Engenharia Civil e Ambiental – Faculdade de Tecnologia – Universidade de Brasília – DF – 155p.
- Tobler, W. R. (1970) A computer movie simulating urban growth in the Detroit region, *Economic Geography*, 46, 234-240.