

## O Uso do Transporte Coletivo & a Oferta de Empregos no Município de São Paulo

*The Use of Collective Transport & The Job Offer in The City of São Paulo (São Paulo State, Brazil)*

*Bárbara M. Giacom Ribeiro*<sup>1</sup>

---

<sup>1</sup> Programa de Pós-Graduação em Sensoriamento Remoto, Instituto Nacional de Pesquisas Espaciais – INPE. Caixa Postal 515 - 12245-970 São José dos Campos - SP, Brasil. barbara@dpi.inpe.br

**Abstract** *The improvement of the obtention and interpretation process of several social reality indices had been the goal of many analyses along the years. The employment of specific statistical tools, added to spatial analysis techniques, in geoprocessing studies, significantly contributes to improve the the quality of results. In this work, spatial statistics is applied to explore the relationship between the daily displacement in São Paulo city and the number of formal jobs, looking forward to better understand one of the determinant factors of mobility conditions in the city.*

**Key words** *Spatial analysis; spatial statistics; geographic information systems; urban mobility.*

**Resumo** *O aprimoramento da obtenção e interpretação dos diversos indicadores da realidade social têm sido, ao longo dos anos, objeto de inúmeras análises. A utilização de ferramentas estatísticas específicas, aliadas às técnicas de análise espacial no contexto de estudos de Geoprocessamento, vem contribuir significativamente para melhoria da qualidade dos resultados. Desta forma, neste artigo, faz-se uso de técnicas de estatística espacial para explorar a relação entre os deslocamentos realizados diariamente no município de São Paulo e o número de empregos formais existente, buscando compreender um dos fatores determinantes das condições de mobilidade no município.*

**Palavras-chave** *Análise espacial; estatística espacial; sistemas de informações geográficas; mobilidade urbana.*

## Introdução

Nas últimas décadas, a urbanização brasileira tem se caracterizado essencialmente por expansões desordenadas, resultantes de um processo de especulação econômica e de exclusão social. Vastas extensões do território têm sido ocupadas e parceladas sem qualquer planejamento e, como consequência, verifica-se a piora da qualidade de vida nos mais diversos aspectos, o que atinge especialmente a população de baixa renda.

O crescimento e a redistribuição populacional, em especial a intra-urbana, proporciona um impacto direto sobre o desenvolvimento das estruturas espaciais da cidade, redistribuindo grupos sócio-econômicos e alterando padrões de uso das áreas urbanas. Por esta razão, a compreensão de fatores que regem esta dinâmica torna-se uma questão essencial para a estimativa de futuras mudanças na cidade, para a consolidação de políticas efetivas no volume e territorialização dos investimentos urbanos bem como a contenção ou estímulo à intensificação de determinados usos.

As transformações urbanas, observadas na Região Metropolitana de São Paulo, iniciadas na década de 1980, podem ser consideradas uma decorrência de políticas econômicas globais (principalmente abertura e desregulamentação da economia), as quais, afetando o desempenho da economia e as condições socioeconômicas da população, interferiram, direta ou indiretamente, na conformação física e no papel das cidades. Estas transformações, ainda em andamento, produziram uma nova espacialidade da desigualdade na cidade de São Paulo. Como um dos principais impactos físicos na estrutura urbana, registra-se um aumento no “fosso” existente entre áreas onde as atividades têm se instalado e as áreas periféricas em geral de moradia (TASCHNER e BOGUS, 2000). O fosso espacial existente entre o local de moradia e atividades socioeconômicas, somado à baixa mobilidade destas populações (falta de acesso às atividades e infra-estruturas públicas), acaba por agravar mais o quadro de desigualdade social.

A desigualdade social e espacialidade periférica destas populações podem ser mitigadas por uma política de transporte público que promova a mobilidade destas pessoas, aumentando assim o acesso às atividades socioeconômicas, à infra-estrutura pública e a serviços essenciais.

Este trabalho busca, por meio do estudo ainda inicial da mobilidade em São Paulo (dados de viagens diárias por modo de deslocamento), indicar algumas metodologias estatísticas de análise de dados espaciais agregados em áreas (Distritos). Este tipo de metodologia permite, a identificação de *clusters* (agrupamentos de áreas semelhantes) e *outliers* (áreas fora do padrão da região) e podem ser utilizadas em análises de quaisquer dados agregados em área e distribuídos espacialmente.

A análise de dados territorializados é influenciada diretamente pela definição dos grupos de pesquisa da amostra ou espacialmente condicionados pela definição das fronteiras das subdivisões territoriais a que estes grupos de indivíduos estiverem associados. A *dependência espacial* é uma característica inerente à representação de dados através de subdivisões territoriais, que pode ser traduzida na tendência de que o valor de uma variável associada a uma dada localização assemelhe-se mais ao valor de suas amostras vizinhas do que ao restante das localizações do conjunto amostral (RAMOS, 2002).

A necessidade de quantificação da dependência espacial presente num conjunto de geodados levou ao desenvolvimento da *estatística espacial*, cuja característica principal é seu foco em inquirir padrões espaciais de lugares e valores, identificando a associação espacial existente entre eles e a variação sistemática do fenômeno por localização (ANSELIN, 1996).

A população, as atividades econômicas e as facilidades de transportes mudam seu estado, suas configurações e seus valores de atributo ao longo do tempo, o que confere um aspecto dinâmico às atividades humanas. Por sua vez, o dinamismo das atividades humanas influencia diretamente na variabilidade espacial e temporal dos deslocamentos urbanos. O conhecimento contínuo desta dinâmica urbana tem papel fundamental na gestão eficiente do tráfego urbano.

Nesse âmbito, o melhor entendimento da crise de mobilidade urbana no Município de São Paulo passa por uma análise mais detalhada das diversas relações entre o uso e a ocupação do solo urbano, os sistemas de transporte e a infra-estrutura viária e a interação entre fator humano, veículo, via pública e meio ambiente.

## Objetivo

Este trabalho buscou estabelecer quais fatores poderiam ser determinantes da quantidade de viagens realizadas diariamente no município de São Paulo, utilizando os modos de transporte coletivos - ônibus, metrô e trem. O enfoque principal, todavia, consiste em explorar a modelagem dessas variáveis utilizando dados da oferta de empregos como variáveis independentes.

## Justificativas

A maior cidade brasileira detém 25% da frota nacional, o que hoje já representa mais de seis milhões de veículos (SMT, 2008). Praticamente, tem-se um carro para cada dois habitantes. A Pesquisa Origem-Destino de 1997 e a Aferição de 2002, realizadas pela Companhia do Metropolitano ("Metrô"), a cada dez anos, desde 1967, abrangendo a área mais fortemente urbanizada da Região Metropolitana de São Paulo (RMSP), identificou, em sua última edição, 30 milhões de deslocamentos diários, sendo que cada terça parte corresponde ao transporte coletivo, individual e aos deslocamentos a pé (Figura 2).

A mobilidade urbana no município de São Paulo é caracterizada por um sistema de transportes complexo, composto de diversas variáveis e subsistemas. Diariamente, movimentam-se na cidade indivíduos oriundos dos vários municípios que formam a RMSP e de localidades além desta, de forma que a questão da mobilidade na cidade engloba, necessariamente, as escalas intra-urbana, metropolitana e regional. Além disso, a cidade é conhecida pela convergência de diversas rodovias estaduais e federais, com destinos outros que não a cidade em si, de forma que sistemas de transporte de carga com destinos e origens diferentes de São Paulo passam obrigatoriamente por dentro do município.

Embora um de seus símbolos seja os enormes congestionamentos de veículos em suas principais vias, o transporte coletivo

representa um papel fundamental no cotidiano da metrópole. São Paulo conta com uma grande estrutura de linhas de ônibus, com uma frota de mais de 15.000 unidades (SMT, 2008). Os trens da CPTM, o Metrô e o sistema de interligação entre eles completam o sistema municipal e estadual de transporte na cidade. Informalmente, também se nota o uso de "lotações" clandestinas para o transporte daqueles que tem pouco acesso ao transporte formal disponível.

Todavia, transitando sobre uma estrutura incapaz de dar vazão ao volume de veículos circulantes, a Secretaria de Estado dos Transportes Metropolitanos (STM) estima que o paulistano perde o equivalente a vinte dias por ano em congestionamentos, avaliados em US\$ 6 milhões por dia em tempo e combustível, sem esquecer que os níveis de poluição gerados pelos meios de transporte atuais respondem entre 73% a 94% dos agentes poluentes da atmosfera (STM, 2005).

A demanda pelo deficiente sistema de transporte no município pode ser traduzida em números com auxílio do Levantamento Censitário de 2000, realizado pelo IBGE, que revelou o agravamento de certas tendências preocupantes no desenvolvimento da RMSP. Dentre aquelas que têm repercussão nas demandas de transporte vale ressaltar a representada pelo desequilíbrio do crescimento demográfico da metrópole: enquanto as zonas centrais dos principais municípios da região, incluindo São Paulo, Guarulhos, ABC e Osasco, perdem população residente no período entre o ano de 1991 e 2000, existem áreas localizadas principalmente nas franjas urbanas que cresceram a taxas extremamente elevadas no mesmo período. A fração mais consolidada da metrópole cai de 6,7 milhões para 6,0 milhões de habitantes, enquanto que as demais áreas crescem de 7,7 milhões para 10,3 milhões. E como a maior quantidade de empregos situa-se nas zonas centrais, fica fácil avaliar as grandes e

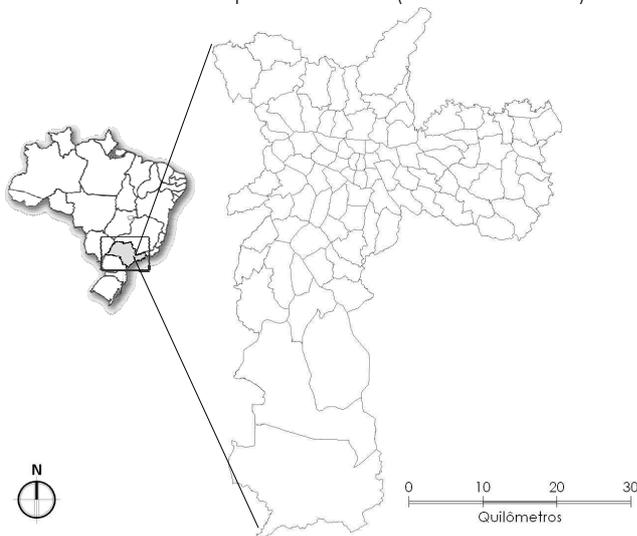
crescentes pressões que se colocarão sobre os serviços de transportes, caso persistam essas tendências (IBGE, 2002).

### Área de Estudo

O município de São Paulo está localizado na Região Metropolitana de São Paulo, com latitude Sul de 23°32'36" e longitude Oeste de 46°37'59". As frações da área de estudo correspondem à divisão do município em seus 96 distritos, conforme a Figura 1 a seguir.

Figura 1

Área de estudo: município de São Paulo (dividido em distritos)



### Materiais

Para a execução do trabalho foram utilizados os softwares TerraView 3.2.1 (INPE, Tecgraf PUC-Rio, FUNCATE, 2008), S-Plus 4.5 (MathSoft Inc., 1998), GeoDA 0.9.5 (Luc Anselin and The Regents of the University of Illinois, 2004).

#### • Banco de Dados

O banco de dados utilizado foi o INFOLOCAL – Sistema de Informação para Planejamento Local, que consiste em uma reunião de informações espaciais georreferenciadas relativas ao município de São Paulo, produzidas pela Secretaria Municipal de Planejamento – SEMPLA-PMSP, a partir de cadastros setoriais públicos e privados. O INFOLOCAL associa a base cartográfica

digital do município de São Paulo *GEOLOG versão 2.1.2* a um conjunto de dados georreferenciados utilizando como ferramenta um *software* de geoprocessamento. Além dos dados originais contidos no banco, foram inseridos e atualizados alguns dados utilizando a Contagem de População de 2002 (IBGE).

Os dados utilizados encontravam-se agregados por Distrito. Dentre todo o conjunto de informações disponível, foram utilizados principalmente os dados relativos às viagens realizadas diariamente no município (divididas em modos de transporte e motivo de deslocamento) e os dados relativos aos postos de emprego (divididos em nove classes: administração pública, comércio, serviços, agropecuária, extração mineral, indústria de transformação, serviços industriais e de utilidade pública, construção civil e outros/ignorado).

A fonte de tais informações é o Cadastro de Empregadores da RAIS (Relação Anual de Informações Sociais), produzido anualmente pelo Ministério do Trabalho. Refere-se ao mercado formal de trabalho, uma vez que devem declarar a RAIS todos os estabelecimentos que tenham mantido, na condição de empregadores, qualquer relação de emprego durante o ano-base do cadastro, bem como todos os demais estabelecimentos legalmente constituídos, mesmo que não tenham mantido vínculos empregatícios ou que tenham mantido trabalhadores avulsos.

Já os dados de tráfego provêm das *Pesquisas de Origem-Destino* realizadas pelo Metrô, que constituem as fontes de informações mais completas sobre os deslocamentos realizados pela população na região Metropolitana de São Paulo. Os dados apresentados são da última O/D, realizada em 1997, atualizados com informações obtidas na *Aferição da Pesquisa Origem-Destino de 2002 (Mini O/D)*, e referem-se às viagens desagregadas por zonas de origem, ou seja, geradas em cada um dos Distritos Municipais.

As viagens diárias com destino a cada Distrito

estão classificadas em seis modos de viagem (coletivo sobre pneus, metrô, ferrovia, individual motorizado e a pé e outros modos de viagem) e em seis motivos de viagens (trabalho, residência, educação, saúde, compras/lazer e outros motivos de viagem).

A Figura 2 mostra os gráficos da divisão das viagens segundo os modos e motivos de deslocamento.

Figura 2

Gráfico A: Número de viagens diárias (em milhões) divididas por modo de transporte para os anos de 1997 e 2002.

Gráfico B: Número de viagens diárias (em milhões) divididas por motivos de deslocamento para os anos de 1997 e 2002.

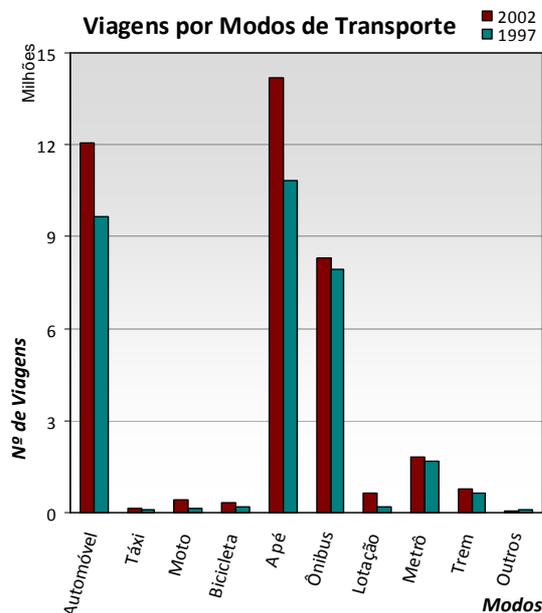


Gráfico A

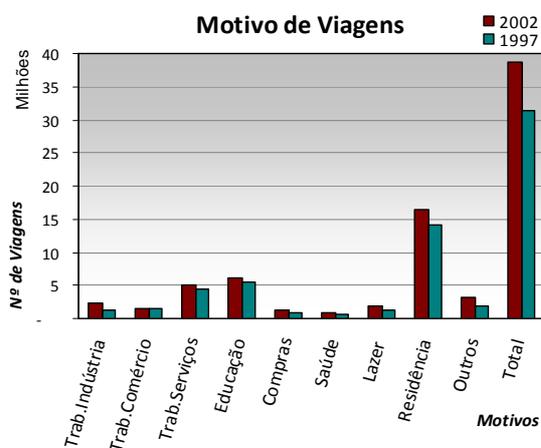


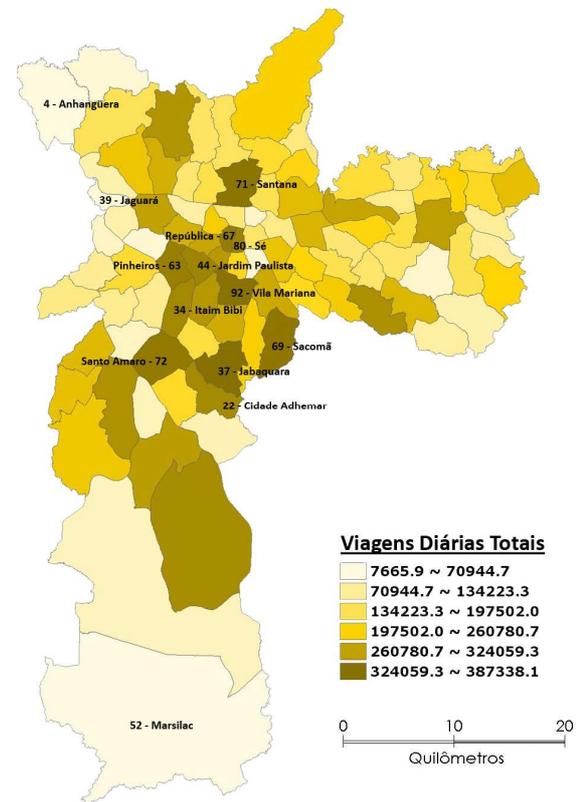
Gráfico B

A Figura 5 exhibe a espacialização desses dados, cuja análise permite afirmar que os distritos que recebem mais viagens diariamente são:

Jabaquara, Vila Mariana, Santana, Sacomã, Santo Amaro, República, Pinheiros, Sé, Jardim Paulista, Itaim Bibi, Cidade Adhemar, todos com mais de 345.000 viagens diárias. No outro extremo, tem-se Marsilac (com 7.266 v.d.), Anhangüera e Jaraguá (com até 40.000 v.d.).

Figura 3

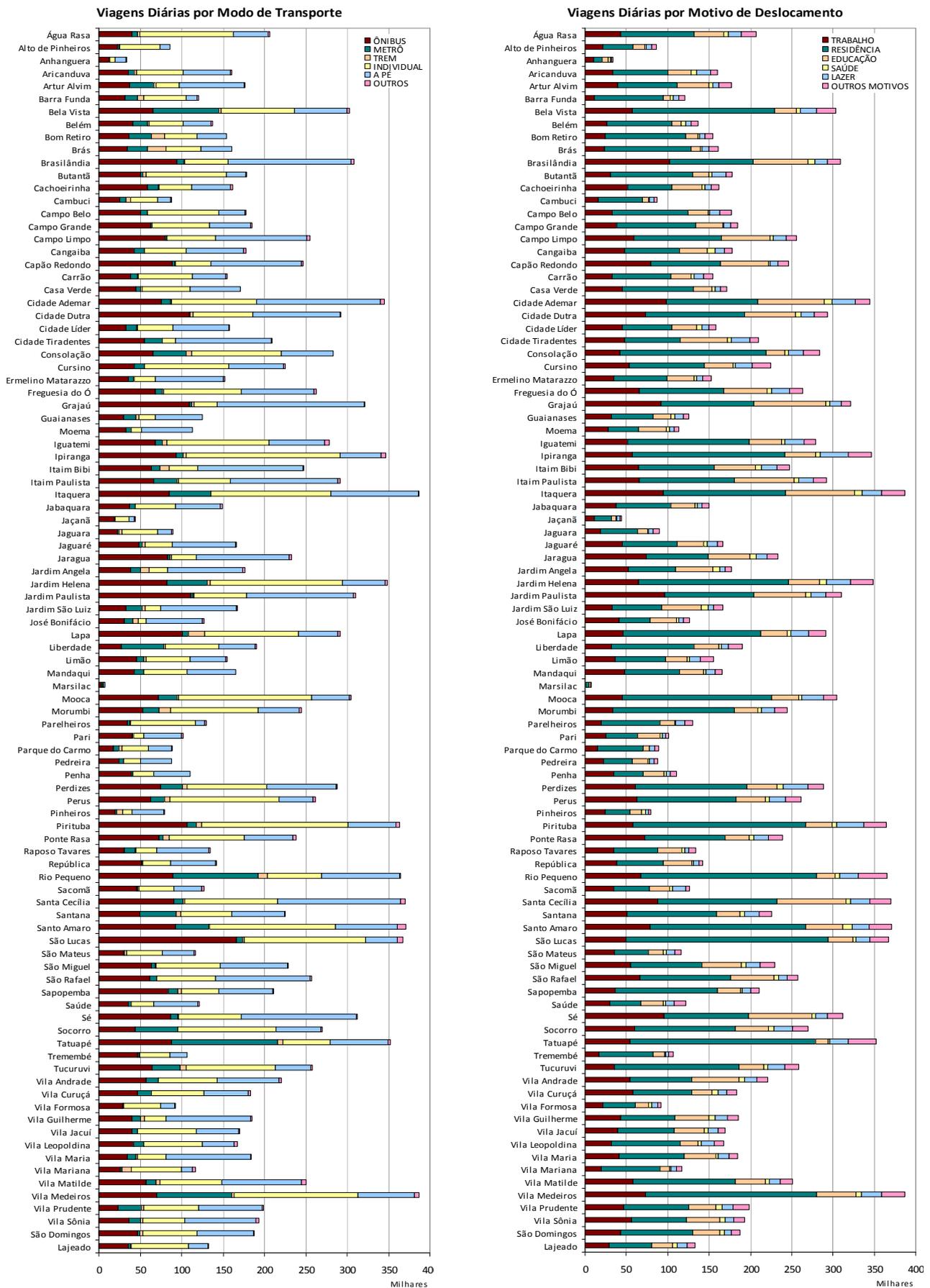
Viagens diárias totais realizadas com destino a cada distrito.



Contudo, distribuição espacial das viagens por modo de transporte ou por motivo de deslocamento não se comporta de maneira semelhante à distribuição do total de viagens diárias. Cada Distrito do município está sujeito a uma série de fatores que condicionam ou limitam a realização das viagens. Estes fatores incluem, por exemplo, a existência das malhas metroviária e ferroviária propriamente ditas, corredores de ônibus, ciclovias. Quando o dado “viagens” é classificado por motivos de deslocamento, faz-se necessário observar a existência de atrativos destes deslocamentos: os Distritos que possuem mais instituições na área da saúde, como hospitais de grande porte, por exemplo, apresentam grande quantidade de viagens por motivos de “Saúde”, já aqueles que

Figura 4

Gráfico A: Número de viagens diárias (em milhares) divididas por modo de transporte para cada Distrito de São Paulo (ano: 2002).  
 Gráfico B: Número de viagens diárias (em milhares) divididas por motivo de deslocamento para cada Distrito de São Paulo (ano: 2002).



possuem maior população, apresentam mais viagens com o motivo de “Residência”.

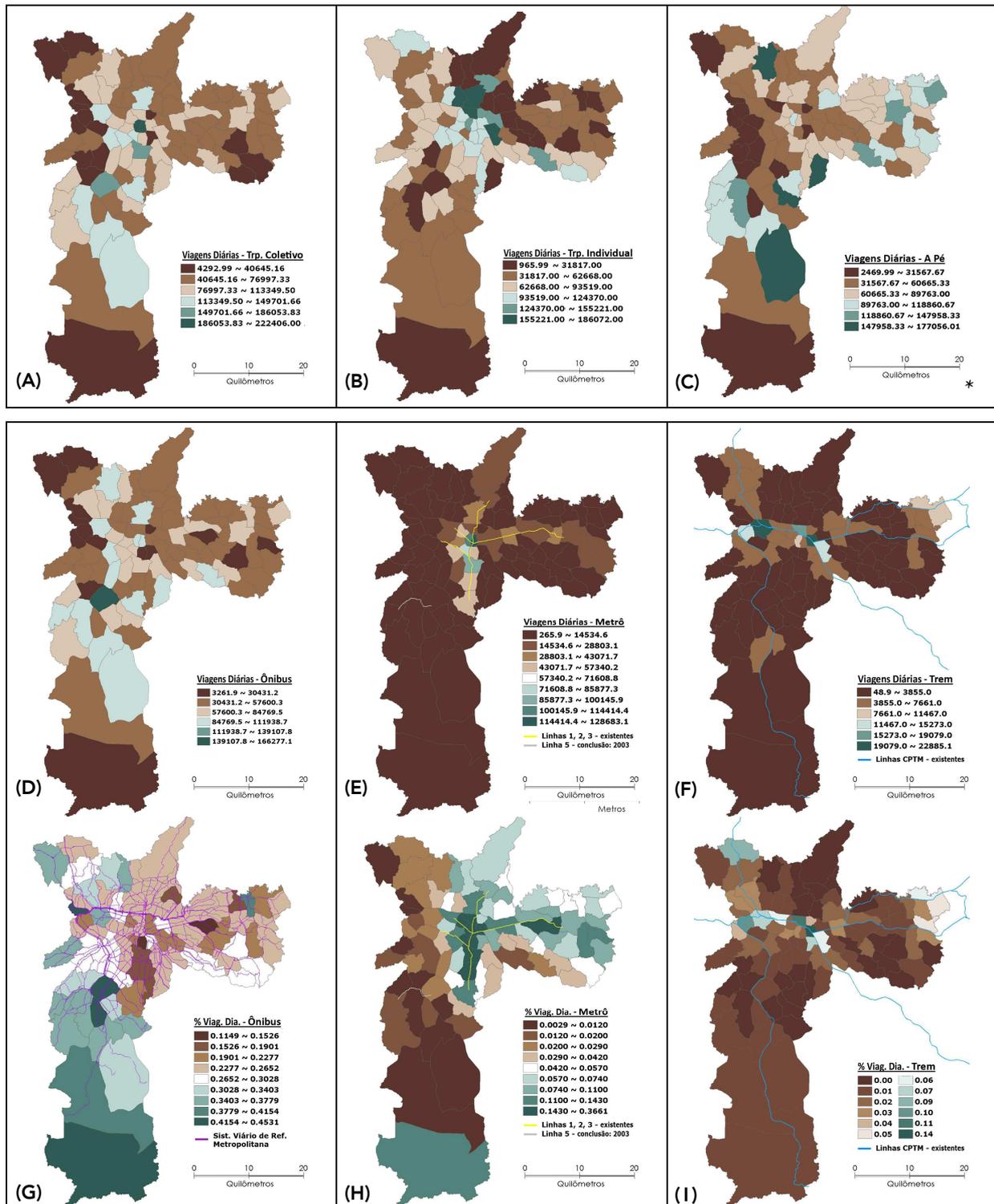
Na Figura 4 são mostrados os gráficos da divisão das viagens para cada Distrito, segundo

os modos e motivos de deslocamento.

A Figura 5 exhibe os mapas cloropléticos com a distribuição espacial das viagens diárias por modo de transporte.

Figura 5

Viagens diárias por modo de transporte: (A) utilizando meios de transporte *Coletivos*, (B) transporte *Individual Motorizado*, (C) deslocamentos *A Pé*. Viagens diárias por modos de transporte *Coletivos*: (D) Ônibus, (E) Metrô, (F) Trem. Porcentagens das viagens diárias, com destino a cada um dos Distritos, realizadas utilizando-se cada modo de transporte: (G) Ônibus, (H) Metrô, (I) Trem.



\* Modo “a pé”: é registrada a viagem realizada a pé da origem ao destino quando a distância percorrida for superior a 500 metros (cerca de cinco quadras). Quando o motivo da viagem é *Trabalho* ou *Educação* a viagem é registrada independente da distância percorrida.

A análise visual dos mapas mostrados na Figura 5 permite dizer, por exemplo, que, apesar de o Distrito de Marsilac (extremo sul do município) receber poucas viagens diárias por ônibus (3.262 – mapa D), esta quantidade representa 42,6% das viagens realizadas por todos os modos de transporte até o local (mapa G). Embora as linhas do Metrô não alcancem a região, 12,6% das viagens realizadas com destino a este Distrito fazem uso de metrô (mapa H), o que sugere que estes deslocamentos utilizam mais de um modo de transporte.

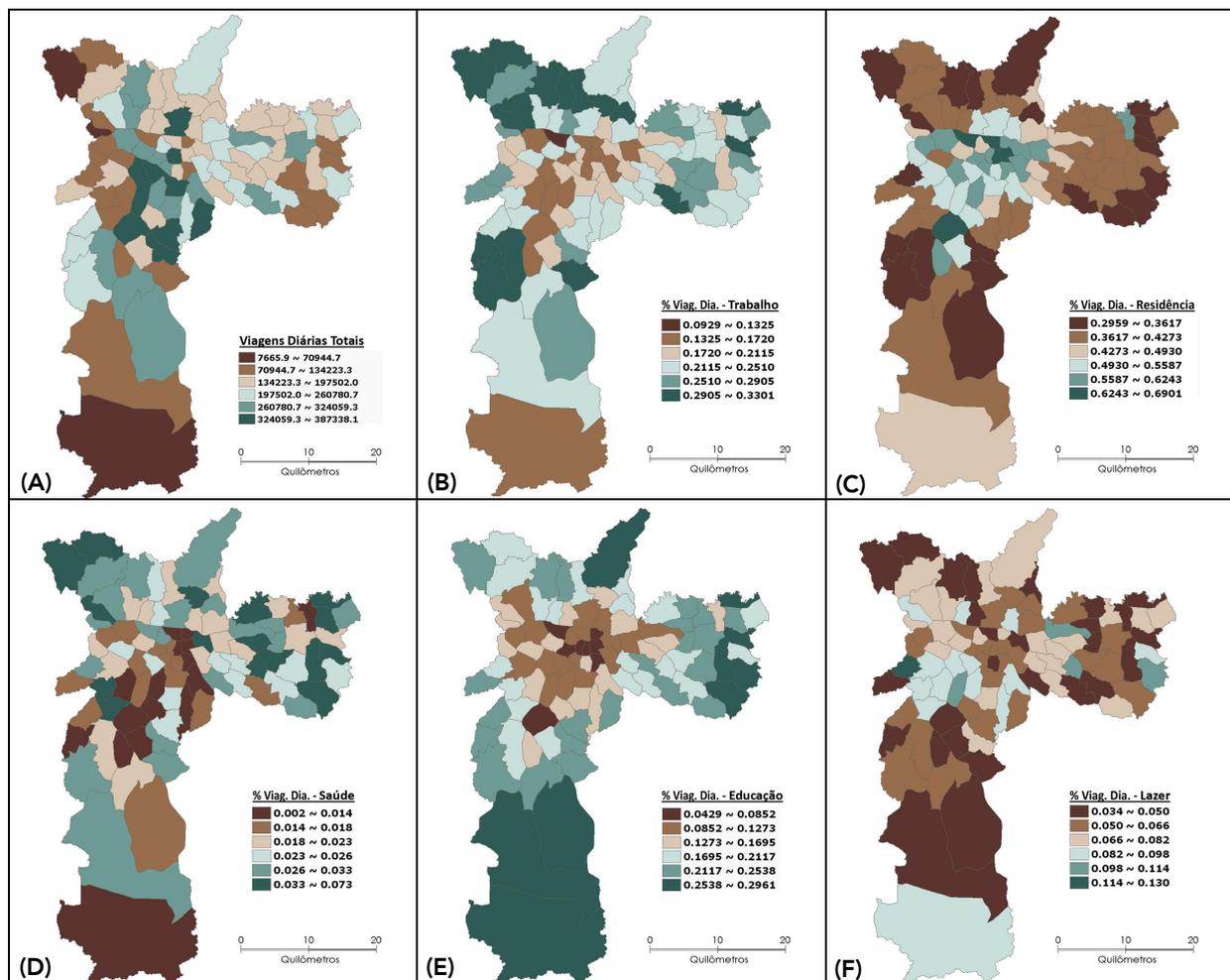
Por sua vez, o Distrito da Sé recebe mais de 350

mil viagens diárias (Figura 3), sendo que 36,6% são por metrô (mapa H), 24,9% por ônibus (mapa G) e apenas 1,7% é realizado em trens (mapa I). Os transportes individuais motorizados (automóveis e motocicletas) respondem por 16,6% das viagens até a Sé.

Na Figura 6 são mostrados os mapas coropléticos com a distribuição espacial das viagens diárias por motivo de transporte, em porcentagens. Cada um dos mapas representa as porcentagens das viagens realizadas por cada motivo em relação ao total de viagens realizadas com destino a cada um dos Distritos.

Figura 6

(A) Viagens diárias totais realizadas para cada Distrito. Os mapas B, C, D, E e F correspondem às porcentagens das viagens diárias totais para cada Distrito realizadas segundo cada motivo de deslocamento: (B) motivo de *Trabalho*, (C) retorno à *Residência*, (D) *Saúde*, (E) *Educação*, (F) *Lazer*.



A análise da Figura 6 permite afirmar também que o Distrito de Marsilac não deve possuir um número significativo de equipamentos de saúde, uma vez que apenas 1,1% das viagens

até este Distrito são realizadas pelo motivo de *Saúde* (mapa D). As viagens por motivo de *Trabalho* também se apresentam pouco representativas do total de viagens realizadas

até o local, correspondendo a 13% (mapa B). A porcentagem relativamente alta das viagens por motivo de *Residência* (45,4% - mapa C) sugere que o Distrito pode possuir a característica de “dormitório”, com taxas muito maiores de habitantes (8.404) do que taxas de emprego (696), educação (837 matrículas em instituições de ensino de todos os níveis) ou atendimentos em saúde.

A Figura 7 exhibe os mapas cloropléticos das viagens diárias com o propósito de *Educação* e

Figura 7

(A) Viagens diárias realizadas com motivo de *Educação*. (B) Matrículas em instituições de ensino públicas e particulares, em todos os níveis, infantil e adulto.

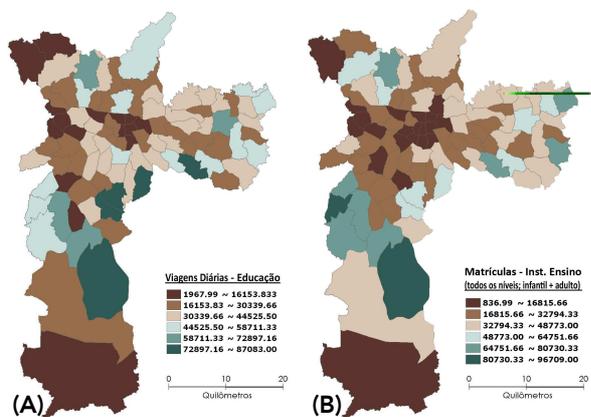


Figura 8

(A) Viagens diárias totais realizadas para cada Distrito. Os mapas B, C, D, E e F correspondem às porcentagens das viagens diárias

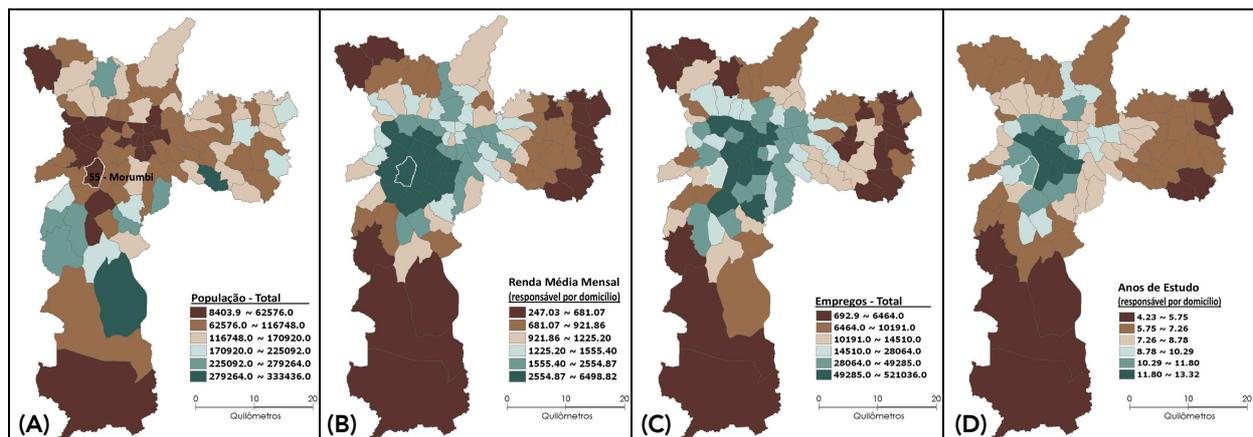
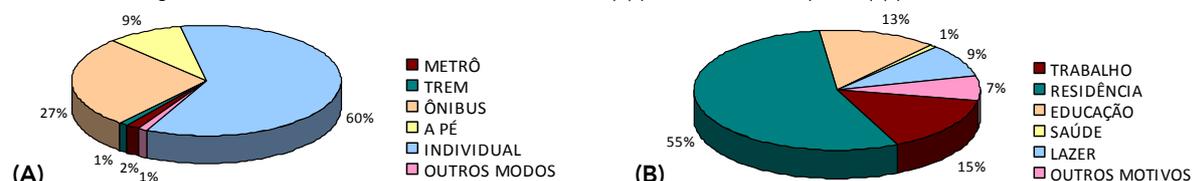


Figura 9

Gráficos das viagens diárias com destino ao Distrito do Morumbi: (A) por modo de transporte, (B) por motivo de deslocamento.



do número de matrículas em instituições de ensino para cada Distrito de São Paulo.

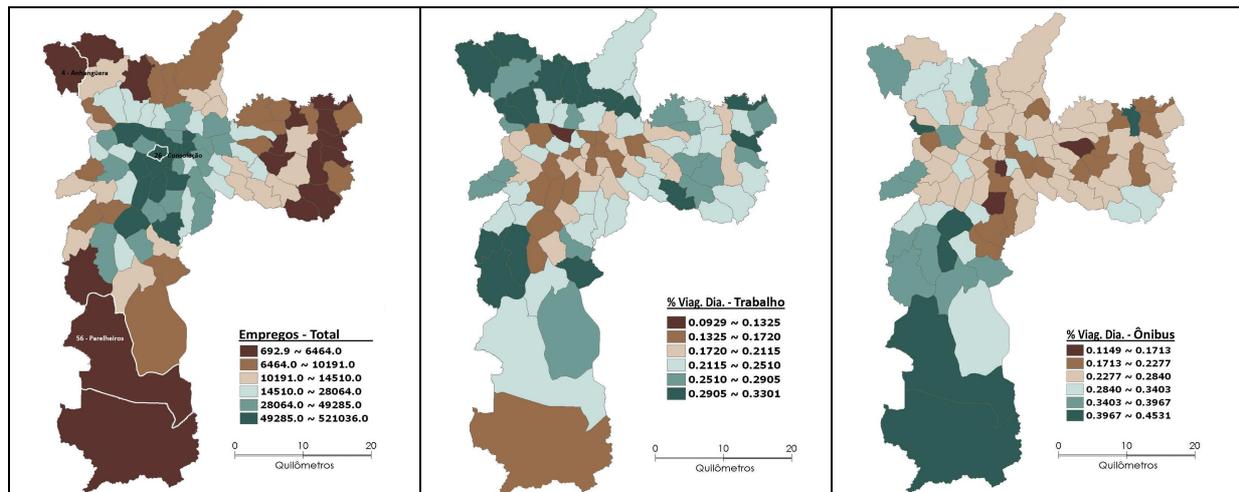
Com a adição de dados suplementares à análise, como população residente nos Distritos, renda mensal e anos de estudo dos responsáveis pelos domicílios, número de empregos, entre outros, novas relações podem ser inferidas dentro do conjunto de dados.

Por exemplo, o fato de o Morumbi possuir baixa população residente (34.568 habitantes), com alta renda média mensal do chefe de família (R\$6.498,82), alta oferta de empregos em geral (cerca de 26.000 postos) e chefes de família com uma média alta de anos de estudo (11,3 anos), permitiria supor uma alta taxa de utilização de automóveis nos deslocamentos diários, modo que, de fato, representa 60% dos deslocamentos diários até o Distrito (Figura 9).

Embora o Distrito possua grande quantidade de empregos em relação à sua população residente, os 55% das viagens diárias totais destinadas ao Morumbi realizadas com motivo de *Residência* podem indicar que boa parte da população que habita o local trabalha em outras localidades.

Figura 10

(A) Viagens diárias totais realizadas para cada Distrito. Os mapas B, C, D, E e F correspondem às porcentagens das viagens diárias



A análise dos mapas mostrados na Figura 10 permite dizer que, na região central do município<sup>1</sup>, é observado maior número de empregos (Consolação: 176.989). Entretanto, constata-se também um baixo percentual de viagens de ônibus chegando ali (22% em relação às viagens por todos os modos). As viagens realizadas diariamente com motivo de *Trabalho* são menores também em direção ao Centro (Consolação: 14,9%), de modo que as pessoas que vão pra lá possuem outros motivos (*Residência, Saúde...*).

Neste cenário, observa-se que quem mora longe da oferta de empregos pode estar deslocando-se mais de ônibus, se comparado a outros modos de transporte (Parelheiros: 39,6%; Anhangüera: 36,2%).

A interpretação dos mapas permite que se estabeleça uma série de relações entre os dados sócio-econômicos disponíveis. Contudo, para melhor traduzir a real relação entre eles, faz-se uso da *estatística espacial*, que busca inquirir padrões espaciais de lugares e valores, identificando a associação espacial existente entre eles e a variação sistemática do fenômeno por localização (ANSELIN, 1996).

<sup>1</sup> O Distrito do Brás apresenta o valor de 521.036 empregos devido à metodologia da declaração dos postos de trabalho: neste Distrito estão "registrados" 489.620 empregos na administração pública que, de fato, distribuem-se em outros Distritos do município.

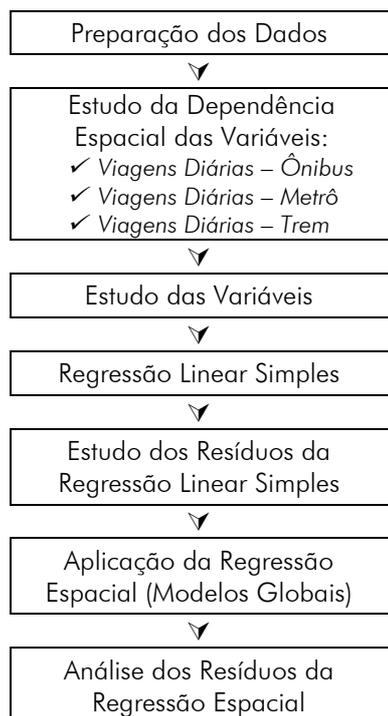
## Metodologia

Este trabalho tem por objetivo utilizar técnicas de análise espacial para aferir a relação entre a quantidade de viagens realizadas diariamente no município de São Paulo, utilizando cada modo de transporte coletivo (ônibus, metrô, trem) e o número de empregos (divididos em empregos na administração pública, comércio, serviços, construção civil, agropecuária, extração mineral, indústria de transformação, e serviços industriais e de utilidade pública), segundo os Distritos do município de São Paulo. Nesse sentido, foram construídos modelos de regressão (simples e múltipla) para análise dos dados. A Figura 11 mostra o fluxo das atividades desenvolvidas.

Usualmente, a análise de regressão visa encontrar um bom ajuste entre os valores preditos por um modelo e os valores observados da variável dependente. Procura também descobrir quais das variáveis explicativas contribuem de forma significativa para o relacionamento linear. A hipótese padrão é que as observações não são correlacionadas e, portanto, os resíduos  $\epsilon_i$  do modelo são independentes e não-correlacionados com a variável dependente, além de apresentar distribuição normal com média zero e variância constante.

Figura 11

Fluxograma das atividades desenvolvidas.



No caso de dados em que exista dependência espacial, é muito pouco provável que a hipótese padrão de observações não correlacionadas seja verdadeira. No caso mais comum, os resíduos continuam apresentando a autocorrelação espacial presente nos dados, que pode se manifestar por diferenças regionais sistemáticas, ou ainda por uma tendência espacial contínua.

Desta forma, a investigação dos resíduos da regressão, em busca de sinais da estrutura espacial, pode fornecer um indicativo da necessidade da utilização de um modelo de regressão espacial. As ferramentas usuais de análise gráfica e o mapeamento dos resíduos podem fornecer a primeira indicação de que os valores observados estão mais correlacionados do que seria esperado sob uma condição de independência (FOTHERINGHAM et al., 2000). Somado à análise gráfica, pode-se utilizar testes estatísticos para verificação de autocorrelação espacial nos resíduos da regressão (por exemplo, a análise do índice  $I$  de Moran).

A análise de regressão em dados espaciais incorpora, na modelagem, a dependência espacial entre os dados, melhorando o poder preditivo do modelo. Primeiramente, faz-se a análise exploratória com o intuito de identificar a estrutura de dependência nos dados, visando a definição da forma de incorporação dessa dependência ao modelo de regressão. Existem dois tipos básicos de modelagem que permitem incorporar o efeito espacial: as de forma global e as de forma local (CÂMARA et al., 2004; ANSELIN, 2002; FOTHERINGHAM et al., 2000).

Os modelos de forma global capturam a estrutura espacial através de um único parâmetro que é adicionado ao modelo de regressão tradicional. Os modelos mais simples, os quais são testados neste estudo, são: Modelo Espacial Auto-Regressivo (*Spatial Auto Regressive - SAR* ou *Spatial Lag Model*) e Modelo do Erro Espacial (*Conditional Auto Regressive - CAR* ou *Spatial Error Model*).

### Análise Exploratória

As técnicas de análise exploratória aplicadas a dados espaciais são essenciais ao desenvolvimento da modelagem em estatística espacial. As Análises Exploratórias de Dados Espaciais (ESDA) consistem em adaptações de ferramentas usuais, sendo consideradas uma extensão da Análise Exploratória de Dados (ANSELIN, 1996). As técnicas de ESDA são apresentadas na forma de índices, que medem a associação espacial (*Índice de Moran*), gráfico de espalhamento (*Moran Scatterplot*) e mapas (*Box Map*); e possuem a função de auxiliar na identificação de agrupamentos de objetos (*clusters*), cujos atributos possuem valores altos e baixos, áreas de transição e casos atípicos (*outliers*). Os elementos básicos das técnicas de ESDA são:

- Matriz de proximidade espacial ( $W$ ): é uma 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. A matriz de proximidade busca estimar os vizinhos de uma determinada área de acordo com uma série de critérios (CÂMARA et al., 2001). No presente estudo, utilizou-se o critério de “contigüidade”, onde polígonos que tocam a área têm peso ‘1’ e os que não tocam têm peso ‘0’.

- Vetor dos desvios ( $Z$ ): é calculado a partir da média geral ( $\mu$ ), e 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. A estimativa da média móvel espacial é obtida pela Equação (1):

$$\hat{\mu}_i = \frac{\sum_{j=1}^n w_{ij} \times y_j}{\sum_{j=1}^n w_{ij}} \quad (1)$$

onde:

$\hat{\mu}_i$  = médias ponderadas;

$\sum_{j=1}^n w_{ij}$  = matriz de proximidade espacial.

### Dependência Espacial

Segundo Câmara et al. (2004), 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 dependem dos valores dessa mesma variável nas localizações vizinhas são a *autocorrelação espacial* e o *variograma*.

O **índice de Moran** é uma ferramenta de medida de autocorrelação espacial que faz essa estimativa a partir do produto dos desvios em relação à média. Formalmente, pode ser traduzido na seguinte fórmula (Equação 2):

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})} \quad (2)$$

onde:

$n$  é o número de áreas;

$z_i$  é o valor do atributo considerado na área  $i$ ;

$\bar{z}$  é o valor médio do atributo na região de estudo;

$w_{ij}$  são os elementos da matriz normalizada de proximidade espacial.

Os valores do Índice de Moran variam de -1 a +1 para dados normalizados, e indicam quão semelhante é cada área analisada em relação à sua vizinhança imediata. Os valores de 0 a 1 indicam correlação direta, enquanto que os valores de 0 a -1, correlação inversa. O valor 0 indica ausência de autocorrelação espacial.

A dependência espacial pode ser visualizada por meio do Diagrama de Espalhamento de Moran (*Moran Scatterplot*). Construído com base nos valores normalizados ( $Z$ ) e comparados com a média dos vizinhos ( $W_z$ ) em um gráfico bidimensional dividido em quatro quadrantes, o Diagrama de Espalhamento permite analisar o comportamento da variabilidade espacial. O índice 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:

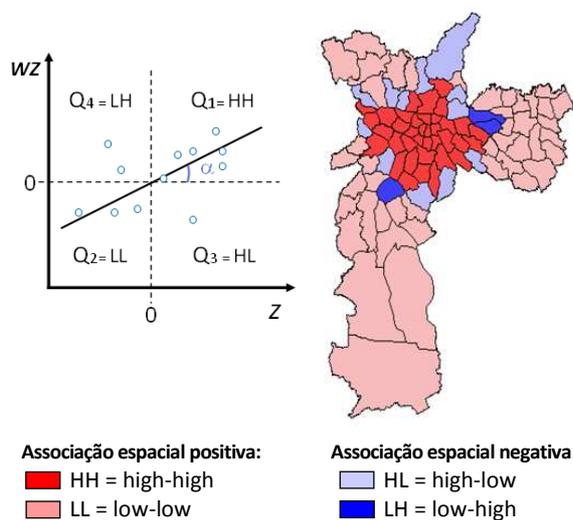
- Q1 (valores positivos e médias positivas) e Q2 (valores negativos e médias negativas): indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes;
- Q3 (valores positivos e médias negativas) e Q4 (valores negativos e médias positivas): indicam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

A Figura 13 mostra um exemplo genérico do Diagrama de Espalhamento de Moran e de seu *Box Map* – uma forma de apresentação do

diagrama de espalhamento que consiste em um mapa temático bidimensional, no qual cada polígono é apresentado indicando-se seu quadrante no diagrama de espalhamento.

Figura 12

Exemplo de Diagrama de Espalhamento de Moran e de um Box Map - mapa temático bidimensional.



O Índice de Moran tem como requisitos básicos a estacionariedade de primeira e segunda ordem. Em caso de violação dessa condição, outra abordagem pode ser utilizada. Uma dessas abordagens é a aplicação de indicadores locais de associação espacial (LISA - *Local Indicators of Spatial Association*).

O **Índice de Moran Local**, que é um LISA, é dado pela Equação 3:

$$I_i(d) = \frac{(x_i - \bar{x})}{s^2} \sum_j w_{ij}(d)(x_j - \bar{x}) \quad (3)$$

onde:

- $w_{ij}(d)$  é a matriz de vizinhança para o par  $i$  e  $j$ ;
- $d$  é a medida de distancia estipulada pelo modelo de vizinhança;
- $\bar{x}$  é a media amostral;
- $x_i$  e  $x_j$  são os valores encontrados em  $i$  e suas vizinhanças  $j$ ;
- $s^2$  é a variância amostral global.

Enquanto os indicadores globais, como o Índice de Moran, fornecem um valor único como medida de associação espacial para todo o conjunto de dados, os indicadores locais

produzem um valor específico para cada área, permitindo, desta forma, a identificação de agrupamentos de áreas com valores de atributos semelhantes (*clusters*), valores anômalos (*outliers*) e de mais de um regime espacial. Segundo Anselin (1996), estes indicadores devem satisfazer duas premissas:

- 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-estacionariedade, e merecem análise detalhada. Autocorrelação com significância superior a 95% indica, a princípio, áreas com grande semelhança em relação a seus vizinhos.

### Regressão Não-Espacial

Um modelo de regressão é uma ferramenta estatística que utiliza o relacionamento existente entre duas ou mais variáveis, de maneira que uma delas possa ser descrita ou o seu valor estimado a partir das demais (BAILEY e GATRELL, 1995).

A relação entre as variáveis pode ser expressa através de uma função matemática que relaciona a variável dependente  $Y$ , a ser explicada, às variáveis independentes  $X$  ( $X_1...X_p$ ), explicativas. A função de regressão de primeira ordem é descrita conforme as Equações 4 e 5:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{ip-1} + \varepsilon \quad (4)$$

$$E(Y) = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{ip-1} \quad (5)$$

onde:

$Y_i$  é a variável dependente;

$\beta_0, \beta_1, \dots, \beta_{p-1}$  são parâmetros multiplicadores das variáveis independentes;

$X_{i1}, \dots, X_{ip-1}$  são variáveis independentes;

$\varepsilon$  é o erro (assume-se que possua distribuição normal, com média zero e desvio padrão  $\sigma^2$ ).

O melhor modelo é alcançado pela técnica dos mínimos quadrados, onde a função linear explicativa paramétrica é minimizada.

Os parâmetros multiplicadores das variáveis independentes ( $\beta_0, \beta_1, \dots, \beta_{p-1}$ ) são calculados segundo a Equação 6:

$$\beta = (X^T X)^{-1} X^T Y \quad (6)$$

onde:

$X^T$  é a matriz transposta de variáveis independentes.

A análise de regressão tem por objetivos encontrar um bom ajuste entre os valores preditos pelo modelo e os valores observados da variável dependente; e descobrir quais das variáveis explicativas contribuem de forma significativa para este relacionamento linear. Para tanto, a hipótese padrão é que as observações não são correlacionadas (as observações necessariamente devem ser independentes), e conseqüentemente, que os resíduos  $\varepsilon_i$  do modelo também são independentes e não-correlacionados com a variável dependente, têm variância constante, e apresentam distribuição normal com média zero.

Entretanto, efeitos espaciais não são considerados na regressão linear simples. Para estes casos, são aplicados modelos de regressão que buscam incorporar a dependência espacial no modelo explicativo. Estes modelos podem tratar a regressão de forma local ou global. Os modelos globais utilizam um único parâmetro para incorporar a estrutura de correlação espacial, enquanto

que os modelos locais utilizam uma regressão espacial para regiões determinadas buscando parâmetros de ajuste locais.

### Regressão Espacial

Em se tratando de dados espaciais, em que pode ocorrer a autocorrelação espacial, a estimativa do modelo de regressão deve incorporar esta estrutura espacial, uma vez que a dependência entre as observações altera o poder explicativo do modelo. A significância dos parâmetros é usualmente superestimada e a existência de variações em larga escala pode até mesmo induzir a presença de associações espúrias (BAILEY e GATRELL, 1995).

Conforme mencionado anteriormente, a inclusão explícita de efeitos espaciais em modelos de regressão pode ser feita de diferentes formas, utilizando modelos de regressão com efeitos espaciais globais ou locais.

O modelo de regressão com efeitos espaciais globais agrega efeitos espaciais globais e supõe que seja possível capturar a estrutura de correlação espacial em um único parâmetro adicionado ao modelo de regressão clássico (BAILEY e GATRELL, 1995; NETER et al., 1996).

Um desses modelos é o *Spatial Lag* (ou SAR - *Spatial Auto Regressive*), que atribui a autocorrelação espacial à variável dependente  $Y$ , ou seja, incorpora o efeito espacial ao modelo. Formalmente esse modelo é expresso pela Equação 7:

$$Y = \rho WY + X\beta + \varepsilon \quad (7)$$

onde:

$W$  é a matriz de proximidade espacial;

$WY$  expressa a dependência espacial em  $Y$ ;

$\rho$  é o coeficiente espacial auto-regressivo.

A hipótese nula para a não existência de autocorrelação é que  $\rho = 0$ . A idéia básica é incorporar a autocorrelação espacial como componente do modelo.

O *Spatial Error* (ou CAR - *Conditional Auto Regressive*) é outro modelo de regressão com

efeitos espaciais globais em que é atribuída a autocorrelação espacial ao erro ( $\varepsilon$ ), ou seja, os efeitos espaciais são considerados ruídos ou perturbações que devem ser removidos (ANSELIN, 1988).

$$Y = X\beta + \varepsilon \quad (8)$$

$$\varepsilon = \lambda W\varepsilon + \xi \quad (9)$$

onde:

$W\varepsilon$  são os erros com efeito espacial;

$\xi$  corresponde aos erros aleatórios com média zero e variância  $\sigma^2$ ;

$\lambda$  é o coeficiente auto-regressivo.

A hipótese nula para a não existência de autocorrelação é que  $\lambda = 0$ , ou seja, o termo de erro não é espacialmente correlacionado.

O modelo de regressão com efeito espacial local é aplicado quando o processo espacial é não-estacionário, e os coeficientes de regressão precisam refletir a heterogeneidade espacial. Nestes casos, pode-se modelar a variação espacial de forma discreta, ao dividir o espaço em sub-regiões estacionárias, chamadas de regimes espaciais, ou então, modelar a tendência espacial de forma contínua, com parâmetros variantes no espaço.

O primeiro caso consiste em dividir a região de estudo em sub-regiões e realizar regressões em separado. O modelo é expresso por meio das Equações 10 e 11:

$$Y_1 = X_1\beta_1 + \varepsilon_1 \quad (10)$$

$$Y_2 = X_2\beta_2 + \varepsilon_2 \quad (11)$$

O segundo caso corresponde à Regressão Ponderada Geograficamente (GWR), e o modelo padrão de regressão é reescrito na forma (Equação 12):

$$Y(s) = \beta(s)X + \varepsilon \quad (12)$$

onde:

$Y(s)$  é a variável aleatória representando o processo no ponto  $s$ ;

$\beta(s)$  é parâmetro multiplicador das variáveis independentes (os parâmetros são estimados no ponto  $s$ );

$(s)$  indica que os parâmetros serão estimados na localização cujas coordenadas são dadas pelo vetor  $s$ ;

$\varepsilon$  é o erro aleatório;

Para estimar os parâmetros deste modelo, a solução padrão por mínimos quadrados para o caso não-espacial, dada conforme a Equação 13, é generalizada utilizando um método de ajuste local (Equação 14):

$$\beta = (X^T X)^{-1} X^T Y \quad (13)$$

$$\beta(s) = (X^T W(s) X)^{-1} X^T W(s) Y \quad (14)$$

onde:

$X^T$  é a matriz transposta de variáveis independentes;

$W(s)$  é a matriz de ponderação das observações.

Os pesos são determinados de modo que as observações próximas ao ponto no espaço onde a estimativa do parâmetro é desejada possuem maior influência no resultado do que as demais observações. Podem ser utilizadas duas funções para o cálculo dos pesos: bi-quadrada e Gaussiana.

Nos modelo globais, cada observação é independente e contribui com a mesma quantidade de informação para a relação em cada ponto (mesmo "peso"). O contexto geográfico dos dados não é utilizado. Por sua vez, a Regressão Ponderada Geograficamente incorpora a informação geográfica no modelo, atribuindo pesos relacionados a distâncias. Através da inferência dos coeficientes de regressão locais, permite estimar as variações espaciais das medidas territorializadas. A tendência espacial é modelada de forma contínua, com parâmetros variantes no espaço. O modelo de regressão é ajustado a cada ponto observado, ponderando todas as demais observações como função da distância a este ponto. São feitos tantos ajustes quantas observações existirem e o resultado será um coeficiente de ajuste para cada localização. Estes coeficientes podem dizer como se comportam espacialmente os relacionamentos entre as variáveis.

No presente trabalho, apenas os modelos que tratam os efeitos espaciais globais foram utilizados. Devido, principalmente, ao grau de agregação dos dados (em Distritos), não foi constatada a necessidade do uso de modelos espaciais locais para melhorar a análise das relações entre as variáveis envolvidas.

### Diagnóstico de Modelos com Efeitos Espaciais

Segundo Câmara et al. (2001), a análise gráfica dos resíduos é o primeiro passo para avaliar a qualidade do ajuste da regressão. Mapear os resíduos consiste numa etapa importante no diagnóstico do modelo, buscando indícios de ruptura dos pressupostos de independência. Uma alta concentração de resíduos positivos (ou negativos) numa parte do mapa é um bom indicador da presença de autocorrelação espacial. Para um teste quantitativo, o procedimento mais comum é utilizar o Índice de Moran sobre os resíduos.

Uma vez que os estimadores e os diagnósticos tradicionais de regressão não consideram os efeitos espaciais, as inferências serão incorretas (as indicações de qualidade de ajuste baseadas em  $R^2$  - coeficiente de determinação<sup>2</sup>, por exemplo). Para se comparar um ajuste obtido por um modelo de regressão padrão com um ajuste obtido por um dos modelos cuja especificação considera a autocorrelação espacial, uma medida como o  $R^2$  não é confiável.

Usualmente a comparação de modelos utiliza o logaritmo da máxima verossimilhança, pois é o que possui melhor ajuste para os dados observados. O critério de informação de Akaike (AIC) é expresso pela Equação 15:

$$AIC = -2 * LIK + 2k \quad (15)$$

onde:

LIK é o log de verossimilhança maximizado;

k é o número de coeficientes de regressão.

Segundo o critério, o melhor modelo é o que possui menor valor de AIC.

Existem outros critérios de informação disponíveis, sendo que a maior parte são variações do AIC, com mudanças na forma de penalização de parâmetros ou observações.

A ANOVA (Análise de Variância) é um importante índice de análise do modelo de Regressão Ponderada Geograficamente. Este indicador compara o modelo de regressão múltipla e o modelo GWR. A hipótese nula é aquela cujo GWR atribui uma melhora significativa ao processo de regressão espacial.

### Resultados

A análise dos dados de viagens diárias por modos de transporte foi realizada, seguindo os mesmos procedimentos metodológicos, para três modos de transporte: ônibus, metrô e trem.

#### 1. Análise de Variáveis

Inicialmente as análises estatísticas buscaram determinar a correlação entre as variáveis explicativas e as independentes.

O processo de geração e validação do modelo de regressão múltipla depende da seleção das variáveis independentes a serem incluídas no modelo. Para tal, pode-se realizar análises a partir de diversos tipos de testes. Neste trabalho, foi utilizado o método da análise da Matriz de Correlação (não tendo sido possível a aplicação do teste *Forward Stepwise*).

Os problemas relacionados à seleção de variáveis e à especificação funcional da função de regressão estão interligados. É importante considerar duas questões na formulação do modelo de regressão: quais as variáveis e em quais formas as mesmas devem ser incluídas no modelo (podem ser originais –  $X_i$ , ou transformadas – por exemplo:  $X^2$ ,  $\text{Log}X$ , etc). Primeiramente, são determinadas as variáveis que serão incluídas na função de regressão, posteriormente, são investigadas as exatas formas nas quais as variáveis serão incluídas na

<sup>2</sup> O coeficiente de determinação ( $R^2$ ) é uma medida da proporção da variabilidade em uma variável que é explicada pela variabilidade de outra variável.

função (CHATERJEE e PRICE, 1977).

Nas Tabelas de 1 a 3 a seguir, são mostradas as matrizes de correlação das variáveis independentes para cada uma das variáveis respostas. Os valores em azul destacam os melhores valores destas correlações para cada caso. Os valores em cinza indicam correlação entre as variáveis independentes, que não devem ser inseridas juntas no modelo, pois podem remover a significância uma da outra.

A **Matriz de Correlação** fornece informações básicas sobre a natureza dos dados analisados, indicando o grau de relação entre as variáveis

independentes e a dependente, assim como o grau de correlação das independentes entre si.

Observando-se a matriz de correlação relativa às *Viagens Diárias de Ônibus* (Tabela 1), verifica-se que a melhor relação entre a variável independente VD.ÔNIBUS e as variáveis dependentes ocorreu para o número de empregos no setor de comércio (EMP.COMÉRCIO): 56,3%. Na Tabela 1 também foram relacionadas outras variáveis independentes que obtiveram os maiores valores de relação com a variável resposta, entretanto, todas se mantiveram abaixo dos 50%.

Tabela 1 - 5

Matrizes de correlação entre as variáveis independentes (população, matrículas em inst. de ensino, mortes no trânsito, média de anos de estudo do responsável por domicílio, empregos nos setores de administração pública, serviços, comércio, construção civil, indústria de transformação e número total de empregos) e cada variável dependente: (1) viagens diárias de ônibus; (2) de metrô; (3) de trem.

(1)	VD.ÔNIBUS	POPULAÇÃO	MATRÍCULAS	MORT.TRÂNS	EMP.SERVIC	EMP.COMÉRC	EMP.CONST.CIV
VD.ÔNIBUS	1	0,4750337	0,4862735	0,4008713	0,4858510	0,5633485	0,3251647
POPULAÇÃO	0,4750337	1	0,9478412	0,8719929	-0,2495215	-0,2967777	-0,2264662
MATRÍCULAS	0,4862735	0,9478412	1	0,8302572	-0,3036648	-0,3054210	-0,2625807
MORTES TRÂNSITO	0,4008713	0,8719929	0,8302572	1	-0,2947588	-0,2714283	-0,2509195
EMP.SERVIÇOS	0,4858510	-0,2495215	-0,3036648	-0,2947588	1	0,7849989	0,7035158
EMP.COMÉRCIO	0,5633485	-0,2967777	-0,3054210	-0,2714283	0,7849989	1	0,6577746
EMP.CONST.CIV	0,3251647	-0,2264662	-0,2625807	-0,2509195	0,7035158	0,6577746	1

(2)	VD.METRÔ	ANOS-EST. RESP	EMP.SERV	EMP.COM
VD.METRÔ	1	0,3777247	0,6232191	0,4389825
ANOS-EST. RESP	0,3777247	1	0,7408163	0,6840262
EMP.SERVIÇOS	0,6232191	0,7408163	1	0,7849989
EMP.COMÉRCIO	0,4389825	0,6840262	0,7849989	1

(3)	VD.TREM	EMP.ADM.PUB	EMP.IND.TR.	EMP.COM	EMP.TOTAL
VD.TREM	1	0,5063546	0,4086121	0,3873036	0,5585145
EMP.ADM.PUBL	0,5063546	1	0,1791721	0,2019911	0,9076402
EMP.IND.TRANSF	0,4086121	0,1791721	1	0,5858017	0,3752309
EMP.COMÉRCIO	0,3873036	0,2019911	0,5858017	1	0,5530292
EMP.TOTAL	0,5585145	0,9076402	0,3752309	0,5530292	1

No caso das *Viagens Diárias de Metrô*, a variável independente com melhor grau de correlação foi a EMP.SERVIÇOS (0,623). As demais variáveis não foram inseridas no modelo relativo aos dados de viagens por metrô (Tabela 2)

As *Viagens Diárias de Trem* demonstraram os maiores graus de relacionamento com as variáveis EMP.ADM.PÚBL. e EMP.TOTAL (50% e 55%, respectivamente) (Tabela 3). Entretanto, pelo fato de os dados relativos aos empregos na administração pública

apresentarem uma “super-agregação” no Distrito do Brás<sup>3</sup>, essas duas variáveis explicativas foram retiradas da análise. Desta forma, não foi possível prosseguir com os estudos acerca das *Viagens Diárias de Trem*

Na matriz de correção da Tabela 1, relativa à variável VD.ÔNIBUS, as variáveis

<sup>3</sup> O Distrito do Brás apresenta o valor de 521.036 empregos devido à metodologia da declaração dos postos de trabalho: neste Distrito estão “registrados” 489.620 empregos na administração pública que, de fato, distribuem-se em outros Distritos do município.

independentes relativas aos dados sócio-econômicos *População e Número de Matrículas* também apresentaram uma relação razoável com a variável resposta, sem, contudo, possuir correlação com as variáveis de emprego envolvidas, de modo que puderam ser inseridas no modelo das *Viagens Diárias por Ônibus*.

Tabela 6

Relação de modelos desenvolvidos para análise das variáveis relativas às viagens de ônibus, metrô e individual.

Modelo 1			
V.DEP:	VD.ÔNIBUS	V.INDEP:	EMP.COMÉRCIO
		V.INDEP:	POPULAÇÃO*
		V.INDEP:	MATRÍCULAS*
Modelo 2			
V.DEP:	VD.ÔNIBUS	V.INDEP:	EMP.SERVIÇOS
		V.INDEP:	POPULAÇÃO*
		V.INDEP:	MATRÍCULAS*
Modelo 3			
V.DEP:	VD.METRÔ	V.INDEP:	EMP.SERVIÇOS

\* a adição das variáveis POPULAÇÃO e MATRÍCULAS (uma de cada vez) nos modelos 1 e 2 causou “remoção” da significância da primeira variável independente de cada modelo (EMP.COMÉRCIO e EMP.SERVIÇOS), sendo excluídas, portanto, de ambos os modelos.

Figura 13

Sumário de resultados do teste de normalidade Kolmogorov-Smirnov para as variáveis relativas às viagens de ônibus, metrô e individual.

One sample Kolmogorov-Smirnov Test of Composite Normality	
data: VD.ONIBUS in Analise.pol	
ks = 0.1294, p-value = 0.0004	
alternative hypothesis: True cdf is not the normal distn. with estimated parameters	
sample estimates:	
mean of x	standard deviation of x
53932.11	27245.62
data: VD.METRO in Analise.pol	
ks = 0.2531, p-value = 0	
alternative hypothesis: True cdf is not the normal distn. with estimated parameters	
sample estimates:	
mean of x	standard deviation of x
15968.4	21829.69

É importante salientar que as duas variáveis respostas utilizadas neste trabalho (VD.ÔNIBUS e VD.METRÔ) apresentaram distribuição normal, de acordo com os testes de Kolmogorov-Smirnov realizados no *software* S-Plus 4.5. Os resultados dos testes são

mostrados na Figura 13.

O mesmo teste de normalidade (Kolmogorov-Smirnov) também foi aplicado nas variáveis independentes a serem utilizadas, e todas apresentaram distribuição normal. Os resultados dos testes para as variáveis explicativas EMP.SERVIÇOS, EMP.COMÉRCIO, POPULAÇÃO e MATRÍCULAS são mostrados na Figura 14. A violação da suposição de normalidade pode enviesar os testes de significância, reduzindo a capacidade preditiva do modelo.

Figura 14

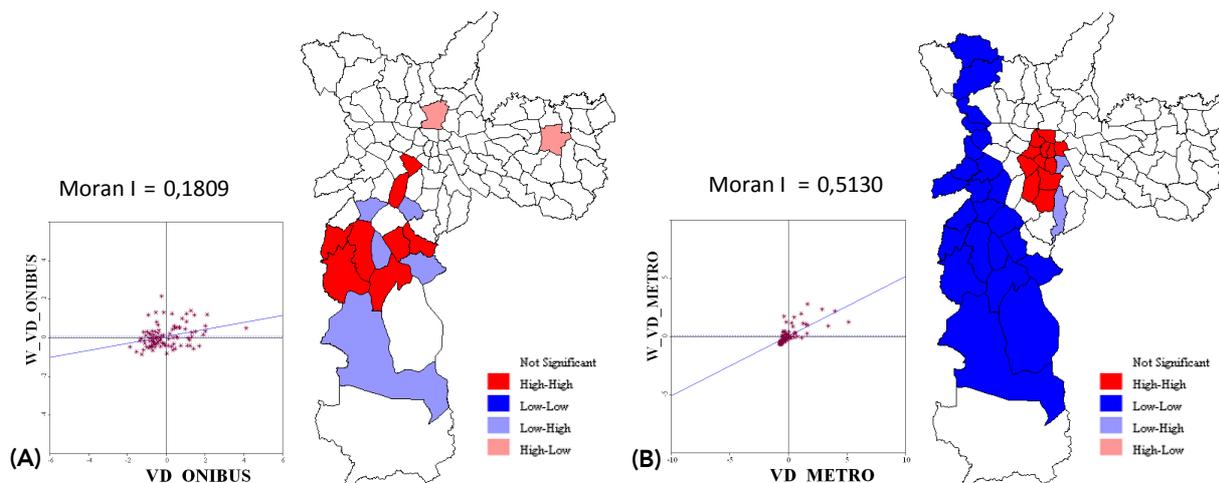
Sumário de resultados do teste de normalidade Kolmogorov-Smirnov para as variáveis independentes EMP.SERVIÇOS, EMP.COMÉRCIO, POPULAÇÃO e MATRÍCULAS.

One sample Kolmogorov-Smirnov Test of Composite Normality	
data: EMP.SERVICOS in Analise.pol	
ks = 0.1764, p-value = 0	
alternative hypothesis: True cdf is not the normal distn. with estimated parameters	
sample estimates:	
mean of x	standard deviation of x
5221.917	4405.128
data: EMP.COMERCIO in Analise.pol	
ks = 0.262, p-value = 0	
alternative hypothesis: True cdf is not the normal distn. with estimated parameters	
sample estimates:	
mean of x	standard deviation of x
12855.85	17240.93
data: POPULACAO in Analise.pol	
ks = 0.128, p-value = 0.0005	
alternative hypothesis: True cdf is not the normal distn. with estimated parameters	
sample estimates:	
mean of x	standard deviation of x
108690.1	64958.16
data: MATRICULAS in Analise.pol	
ks = 0.1051, p-value = 0.0009	
alternative hypothesis: True cdf is not the normal distn. with estimated parameters	
sample estimates:	
mean of x	standard deviation of x
33369.18	20356.64

Em seguida, foram calculados os Índices de Moran para três variáveis respostas. Os valores encontrados para o Índice de Moran Global nos três casos indicaram existência de dependência espacial. A Figura 15 mostra os Diagramas de Espalhamentos obtidos.

Figura 15

Diagramas de Espalhamento de Moran e Moran Map para as variáveis relativas às viagens de ônibus (A) e metrô (B).



Em relação à estrutura espacial da variável VD.ÔNIBUS, o Índice Global de Moran ( $I=0,18$ ) indica a existência de dependência espacial, e o *Moran Map*<sup>4</sup> do Índice de Moran Local (Figura 15A) revela uma presença de regimes espaciais com valores positivos e médias locais positivas em alguns Distritos ao sul do município (cor vermelha), ou seja, *clusters* de alto número de viagens diárias de ônibus. Neste caso, não foram detectados locais de associação espacial positiva *Low-Low*. A maior parte dos Distritos não apresentou valores significativos (cor branca).

Já a estrutura espacial da variável VD.METRÔ, cujo Índice Global de Moran ( $I = 0,51$ ) indica a existência mais forte de dependência espacial, apresentou comportamento conforme o esperado, devido às restrições físicas da malha do Metrô propriamente dita (Figura 15B). O *Moran Map* do Índice de Moran Local revela uma presença forte de regimes espaciais com valores positivos e médias locais positivas nos Distritos centrais município (cor vermelha), ou seja, *clusters* de

alto número de viagens diárias de metrô; e valores negativos e médias locais negativas nos Distritos das áreas sul e extremo oeste (cor azul) – *clusters* de baixo número de viagens diárias de metrô. Neste caso, não foram detectados locais de associação espacial negativa *High-Low*, e apenas dois Distritos com a associação *Low-High* (cor lilás). Mais uma vez, grande parte dos Distritos não apresentou valores significativos (cor branca).

## 2. Regressão Linear Simples

As regressões lineares simples foram realizadas para os três pares de variáveis / modelos (Tabela 6). Os resultados são mostrados nas Figuras 16 e 17.

Através da análise desses resultados, verifica-se que os testes *F* e *t* realizados pra os Modelos 1, 2 e 3 são significativos, considerando significativamente as variáveis *empregos no setor de comércio* (Modelo 1) e *empregos no setor de serviços* (Modelo 2) na modelagem das *viagens diárias de ônibus*; e a variável *empregos no setor de serviços* (Modelo 3) na modelagem das *viagens diárias de metrô*. Observa-se também que os resíduos não possuem distribuição normal. Todavia, enquanto os resíduos dos Modelos 1 e 2 possuem variâncias constantes, o Modelo 3 apresenta resíduos heterocedásticos.

<sup>4</sup> O mapa temático do índice Moran é chamado de "*LISA Cluster Map*" no software GeoDA, que foi utilizado nesta etapa do trabalho. O nome utilizado neste trabalho, "*Moran Map*", foi assim empregado com o intuito de fazer alusão direta ao dado espacializado. Da mesma forma, o *Lisa Map* a que se refere neste trabalho, é chamado de "*LISA Significance Map*" no software GeoDA.

Figura 16

Resultados das Regressões Lineares realizadas para os Modelos 1 e 2 (VD.ÔNIBUS x EMP.COMÉRCIO, VD.ÔNIBUS x EMP.SERVIÇOS) e os respectivos Diagrama de Dispersão.

**Modelo 1**

REGRESSION SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION				
Data set : Analise_pol				
Dependent Variable :	VD_ONIBUS	Number of Observations :	96	
Mean dependent var :	53932.1	Number of Variables :	2	
S.D. dependent var :	27103.3	Degrees of Freedom :	94	
R-squared :	0.317361	F-statistic :	43.701	
Adjusted R-squared :	0.310099	Prob(F-statistic) :	2.29701e-009	
Sum squared residual:	4.81402e+010	Log likelihood :	-1097.8	
Sigma-square :	5.1213e+008	Akaike info criterion :	2199.61	
S.E. of regression :	22630.3	Schwarz criterion :	2204.74	
Sigma-square ML :	5.0146e+008			
S.E of regression ML:	22393.3			
Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	35737.4	3593.045	9.946271	0.0000000
EMP_COM	3.484298	0.5270717	6.610672	0.0000000
REGRESSION DIAGNOSTICS				
MULTICOLLINEARITY CONDITION NUMBER 2.747277				
TEST ON NORMALITY OF ERRORS				
TEST	DF	VALUE	PROB	
Jarque-Bera	2	10.94949	0.0041913	→ NÃO NORMAL
DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	
Breusch-Pagan test	1	1.698719	0.1924556	→ HOMOCEDEÁSTICO
Koenker-Bassett test	1	1.342601	0.2465757	→ HOMOCEDEÁSTICO

**Modelo 2**

REGRESSION SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION				
Data set : Analise_pol				
Dependent Variable :	VD_ONIBUS	Number of Observations :	96	
Mean dependent var :	53932.1	Number of Variables :	2	
S.D. dependent var :	27103.3	Degrees of Freedom :	94	
R-squared :	0.236051	F-statistic :	29.0449	
Adjusted R-squared :	0.227924	Prob(F-statistic) :	5.22568e-007	
Sum squared residual:	5.38743e+010	Log likelihood :	-1103.21	
Sigma-square :	5.7313e+008	Akaike info criterion :	2210.41	
S.E. of regression :	23940.1	Schwarz criterion :	2215.54	
Sigma-square ML :	5.6119e+008			
S.E of regression ML:	23689.5			
Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	44061.59	3053.6	14.42939	0.0000000
E_SERVIC	0.7677841	0.1424637	5.389332	0.0000005
REGRESSION DIAGNOSTICS				
MULTICOLLINEARITY CONDITION NUMBER 1.999317				
TEST ON NORMALITY OF ERRORS				
TEST	DF	VALUE	PROB	
Jarque-Bera	2	19.73246	0.0000519	→ NÃO-NORMAL
DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	
Breusch-Pagan test	1	1.991786	0.1581543	→ HOMOCEDEÁSTICO
Koenker-Bassett test	1	1.355325	0.2443494	→ HOMOCEDEÁSTICO

\* A não-normalidade dos resíduos indica que pode haver agrupamento espacial. A homocedasticidade dos erros indica que possuem variância constante.

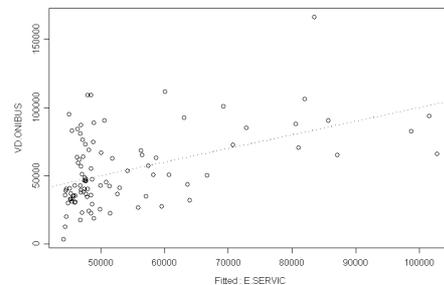
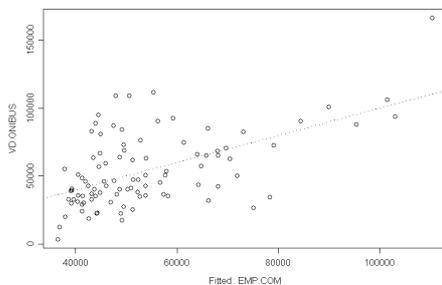
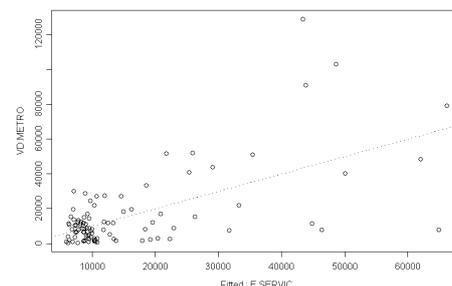


Figura 17

Resultados da Regressão Linear realizada para o Modelo 3 (VD.METRÔ x EMP.SERVIÇOS) e o Diagrama de Dispersão.

**Modelo 3**

REGRESSION SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION				
Data set : Analise_pol				
Dependent Variable :	VD_METRO	Number of Observations :	96	
Mean dependent var :	15968.4	Number of Variables :	2	
S.D. dependent var :	21715.7	Degrees of Freedom :	94	
R-squared :	0.388402	F-statistic :	59.6958	
Adjusted R-squared :	0.381896	Prob(F-statistic) :	1.19292e-011	
Sum squared residual:	2.76875e+010	Log likelihood :	-1071.25	
Sigma-square :	2.94548e+008	Akaike info criterion :	2146.51	
S.E. of regression :	17162.4	Schwarz criterion :	2151.64	
Sigma-square ML :	2.88412e+008			
S.E of regression ML:	16982.7			
Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	5823.965	2189.09	2.660451	0.0091762
E_SERVIC	0.789092	0.1021305	7.726308	0.0000000
REGRESSION DIAGNOSTICS				
MULTICOLLINEARITY CONDITION NUMBER 1.999317				
TEST ON NORMALITY OF ERRORS				
TEST	DF	VALUE	PROB	
Jarque-Bera	2	252.5842	0.0000000	→ NÃO NORMAL
DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	
Breusch-Pagan test	1	113.413	0.0000000	→ HETEROCEDÁSTICO
Koenker-Bassett test	1	23.95176	0.0000010	→ HETEROCEDÁSTICO



**3. Agrupamento Espacial**

O Índice Local de Moran também pode ser calculado de maneira multivariada, associando-se duas variáveis. Na Figura 18 encontram-se o Moran Scatterplot, Moran Map e Lisa Map para os Modelos 1, 2 e 3.

Grosso modo, pode-se dizer que os Moran

Maps multivariados apresentaram-se muito semelhantes aos Moran Maps das variáveis VD.ÔNIBUS e VD.METRÔ (Figura 15).

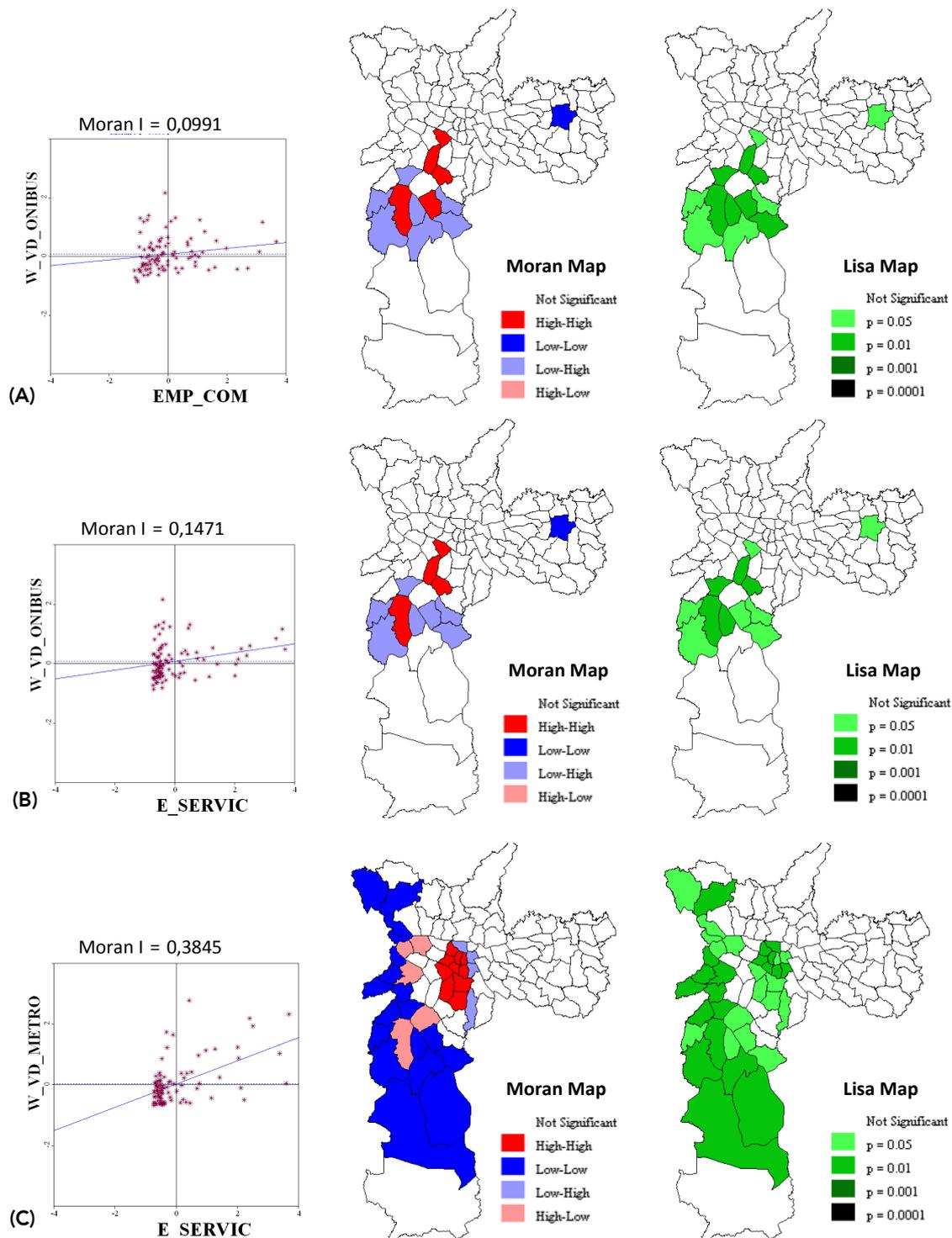
Nos três casos, a presença de clusters de associação espacial positiva (vermelho: High-High, azul: Low-Low) vai ao encontro da hipótese inicial, de que quanto maior a oferta

de empregos (no setor de comércio ou serviços), maior é a quantidade de viagens diárias em ônibus/metrô. Já os *clusters* de associação espacial negativa (lilás: *Low-High*, rosa: *High-Low*) contrapõe-se à hipótese inicial.

A Figura 18C mostra o *Moran Map* das variáveis VD.METRÔ e EMP.SERVIÇOS. O Moran multivariado verifica a influência espacial que o número de empregos no setor de serviços tem sobre a quantidade de viagens diárias de metrô dos distritos vizinhos.

Figura 18

Moran Scatterplot, Moran Map e Lisa Map dos Modelos 1, 2 e 3.



#### 4. Análise dos resíduos de Regressão

A etapa seguinte consistiu na análise da dependência espacial dos resíduos (Figura 18).

No caso dos Modelos 1 e 2, observa-se que o Índice de Moran, apesar não ser alto, é significativo, o que indica correlação espacial dos resíduos. Os testes de dependência espacial por multiplicadores de Lagrange indicam que o melhor modelo a ser aplicado seria o modelo ERROR, uma vez que o teste com o modelo LAG não foi significativo. Entretanto, os dois modelos foram aplicados buscando uma diminuição do AKAIKE.

Figura 18

Resultados do diagnóstico de dependência espacial dos Modelos 1, 2 e 3.

DIAGNOSTICS FOR SPATIAL DEPENDENCE				Modelo 1
FOR WEIGHT MATRIX : pesos.GAL (row-standardized weights)				
TEST	MI/DF	VALUE	PROB	
Moran's I (error)	0.273045	4.3824957	0.0000117	
Lagrange Multiplier (lag)	1	5.5044142	0.0189685	
Robust LM (lag)	1	5.4528457	0.0195365	
Lagrange Multiplier (error)	1	16.1439576	0.0000587	

DIAGNOSTICS FOR SPATIAL DEPENDENCE				Modelo 2
FOR WEIGHT MATRIX : pesos.GAL (row-standardized weights)				
TEST	MI/DF	VALUE	PROB	
Moran's I (error)	0.226285	3.7075578	0.0002093	
Lagrange Multiplier (lag)	1	3.8504373	0.0497330	
Robust LM (lag)	1	9.6826421	0.0018602	
Lagrange Multiplier (error)	1	11.0880144	0.0008689	

DIAGNOSTICS FOR SPATIAL DEPENDENCE				Modelo 3
FOR WEIGHT MATRIX : pesos.GAL (row-standardized weights)				
TEST	MI/DF	VALUE	PROB	
Moran's I (error)	0.453509	7.1711090	0.0000000	
Lagrange Multiplier (lag)	1	32.8458304	0.0000000	
Robust LM (lag)	1	0.0295862	0.8634326	
Lagrange Multiplier (error)	1	44.5362012	0.0000000	

O Modelo 3 apresentou um Índice de Moran mais alto (e significativo), indicando correlação espacial dos resíduos. Ambos os testes de dependência espacial por multiplicadores de Lagrange também mostraram-se significativos, de modo que os dois modelos, LAG e ERROR, foram aplicados para se verificar uma diminuição do valor de AKAIKE.

#### 5. Regressão Espacial

Para os três Modelos (VD.ÔNIBUS × EMP.COMÉRCIO, VD.ÔNIBUS × EMP.SERVIÇOS e VD.METRÔ × EMP.SERVIÇOS) foram aplicados os modelos globais *Spatial Lag* e *Spatial Error*. As Figuras 19, 20 e 21 demonstram os resultados destas regressões.

Nos três casos, a aplicação dos modelos *Spatial Lag* e *Spatial Error* diminuiu os valores de AKAIKE em relação aos valores obtidos com as regressões lineares simples.

Os testes para homocedasticidade apresentaram-se válidos quando da aplicação do modelo *Spatial Lag* para os Modelos 1 e 2 (VD.ÔNIBUS × EMP.COMÉRCIO e VD.ÔNIBUS × EMP.SERVIÇOS); já para o Modelo 3 (VD.METRÔ × EMP.SERVIÇOS), o teste de para heterocedasticidade demonstrou que os resíduos não apresentam variância constante. Com a aplicação do modelo *Spatial Error*, apenas o Modelo 1 obteve resultado válido para homocedasticidade; os testes para os Modelos 2 e 3 indicaram heterocedasticidade dos resíduos.

Entretanto, os coeficientes de regressão obtidos para as três situações, após a aplicação dos modelos espaciais *Lag* e *Error*, não foram significativos (sempre maiores que 0,92). Este fato remove qualquer aspecto positivo da aplicação dos modelos espaciais *Spatial Lag* e *Spatial Error*, uma vez que não assegura nenhuma confiança aos resultados (significâncias:  $\alpha > 92\%$  / desejado:  $\alpha < 5\%$ ).

Para o Modelo 1, a análise dos valores preditos pelos modelos espaciais, em relação aos valores observados (Figuras 19C, 19D, 20C e 20D) mostra que os modelos não foram capazes de modelar a relação entre as variáveis resposta e explicativa, ou seja, as *viagens diárias realizadas pelo modo de transporte ônibus* não podem ser modeladas utilizando-se somente a variável independente *empregos no setor de comércio*. Observando-se os mapas cloropléticos de resíduos (*Lag* e *Error*), é evidente que os resíduos ultrapassaram um limite aceitável, ao extrapolar os próprios valores observados originalmente na variável VD.ÔNIBUS (Figuras 19E e 20E).

O mesmo ocorreu com o Modelo 2: a análise dos valores preditos pelos modelos espaciais, em relação aos valores observados (Figuras 21C, 21D, 22C e 22D) mostra que os modelos

Figura 19

(A) Resultados da aplicação do *Spatial Lag Model* no Modelo 1 (VD.ÔNIBUS x EMP.COMÉRCIO), (B) gráfico relacionando os valores originais e preditos e mapas coropléticos mostrando: (C) número original de viagens diárias de ônibus, (D) valores de VD.ÔNIBUS preditos e (E) resíduos.

REGRESSION - MAXIMUM LIKELIHOOD ESTIMATION		SUMMARY OF OUTPUT: SPATIAL LAG MODEL	
Data set	: Análise_pol	Number of Observations	: 96
Spatial Weight	: pesos_GAL	Number of Variables	: 3
Dependent Variable	: VD_ONIBUS	Degrees of Freedom	: 93
Mean dependent var	: 53932.1		
S.D. dependent var	: 27103.3		
Lag coeff. (Rho)	: 0.266185		
R-squared	: 0.361317 ↑	Log likelihood	: -1095.38
Sq. Correlation	: -	Akaike info criterion	: 2196.75 ↓
Sigma-square	: 4.69171e+008	Schwarz criterion	: 2204.44
S.E of regression	: 21660.4		

Variable	Coefficient	Std. Error	z-value	Probability
W_VD_ONIBUS	0.266185	8.636334	0.03082152	0.9754118
CONSTANT	21759.94	0.0004523451	4.810474e+007	0.0000000
EMP_COM	3.321107	1.982233	1.675437	0.0938484

REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST		DF	VALUE	PROB
Breusch-Pagan test		1	2.368802	0.1237820

REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR SPATIAL DEPENDENCE				
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : pesos_GAL				
TEST		DF	VALUE	PROB
Likelihood Ratio Test		1	4.856278	0.0275457

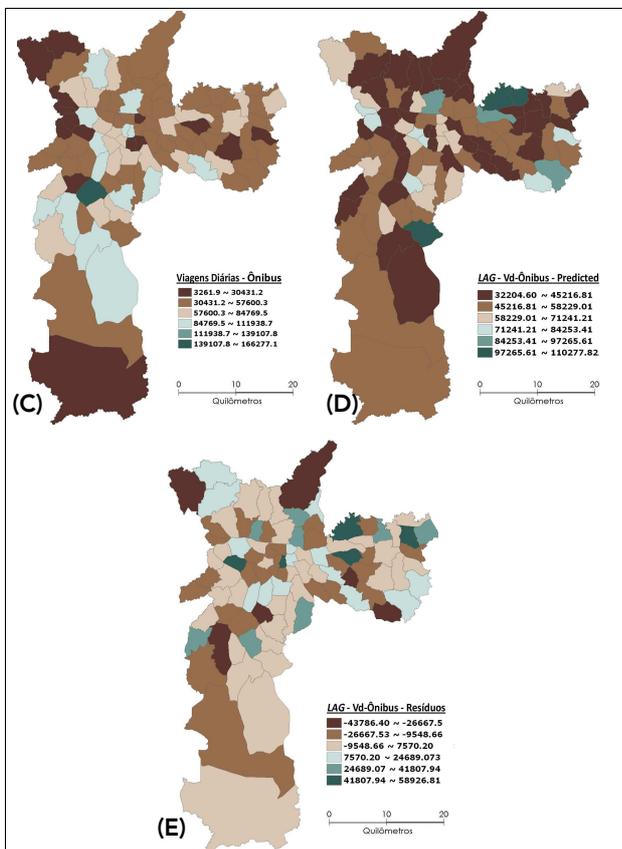
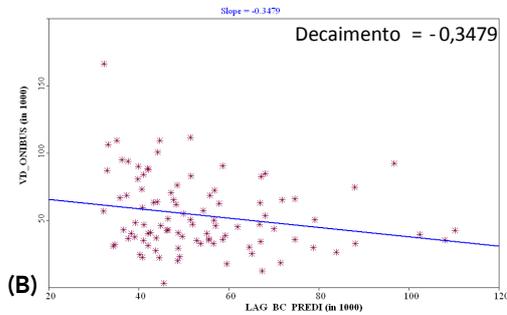


Figura 20

(A) Resultados da aplicação do *Spatial Error Model* no Modelo 1 (VD.ÔNIBUS x EMP.COMÉRCIO), (B) gráfico relacionando os valores originais e preditos e mapas coropléticos mostrando: (C) número original de viagens diárias de ônibus, (D) valores de VD.ÔNIBUS preditos e (E) resíduos.

REGRESSION - MAXIMUM LIKELIHOOD ESTIMATION		SUMMARY OF OUTPUT: SPATIAL ERROR MODEL	
Data set	: Análise_pol	Number of Observations	: 96
Spatial Weight	: pesos_GAL	Number of Variables	: 2
Dependent Variable	: VD_ONIBUS	Degree of Freedom	: 94
Mean dependent var	: 53932.114583		
S.D. dependent var	: 27103.347378		
Lag coeff. (Lambda)	: 0.535295		
R-squared	: 0.462730 ↑	R-squared (BUSE)	: -
Sq. Correlation	: -	Log likelihood	: -1089.802464
Sigma-square	: 394673885.068080	Akaike info criterion	: 2183.6 ↓
S.E of regression	: 19866.4	Schwarz criterion	: 2188.733624

Variable	Coefficient	Std. Error	z-value	Probability
CONSTANT	30343.1	0.0002291892	1.323932e+008	0.0000000
EMP_COM	4.399088	2.148103	2.047894	0.0405703
LAMBDA	0.5352946	9.669383	0.05535975	0.9558517

REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST		DF	VALUE	PROB
Breusch-Pagan test		1	1.173081	0.2787690

REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR SPATIAL DEPENDENCE				
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : pesos_GAL				
TEST		DF	VALUE	PROB
Likelihood Ratio Test		1	16.00265	0.0006633

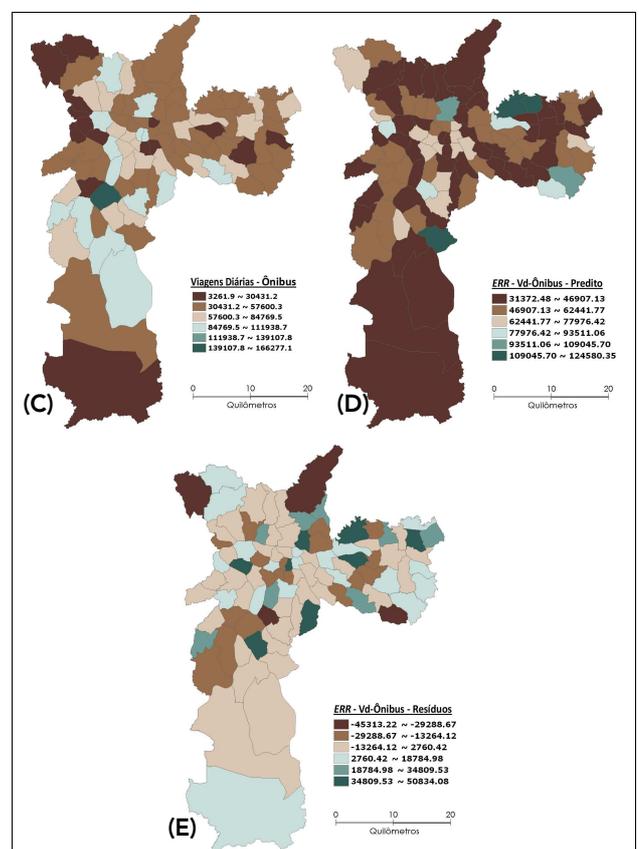
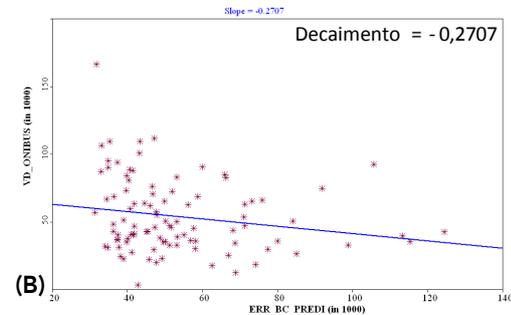


Figura 21

(A) Resultados da aplicação do *Spatial Lag Model* no Modelo 2 (VD.ÔNIBUS x EMP.SERVIÇOS), (B) gráfico relacionando os valores originais e preditos e mapas cloropléticos mostrando: (C) número original de viagens diárias de ônibus, (D) valores de VD.ÔNIBUS preditos e (E) resíduos.

REGRESSION		SUMMARY OF OUTPUT: SPATIAL LAG MODEL		
<b>- MAXIMUM LIKELIHOOD ESTIMATION</b>				
Data set	: Análise pol			
Spatial Weight	: pesos.GAL			
Dependent Variable	: VD_ONIBUS	Number of Observations	: 96	
Mean dependent var	: 53932.1	Number of Variables	: 3	
S.D. dependent var	: 27103.3	Degrees of Freedom	: 93	
Lag coeff. (Rho)	: 0.266185			
R-squared	: 0.270118 ↑	Log likelihood	: -1101.58	
Sq. Correlation	: -	Akaike info criterion	: 2209.16 ↓	
Sigma-square	: 5.36165e+008	Schwarz criterion	: 2216.85	
S.E of regression	: 23155.2			
Variable	Coefficient	Std. Error	z-value	Probability
W_VD_ONIBUS	0.2298911	7.809006	0.02943923	0.9765142
CONSTANT	31940.96	0.0004231422	7.548517e+007	0.0000000
E_SERVIC	0.7143456	7.257269	0.09843174	0.9215894
REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	
Breusch-Pagan test	1	2.93996	0.0864129	
REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR SPATIAL DEPENDENCE				
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : pesos.GAL				
TEST	DF	VALUE	PROB	
Likelihood Ratio Test	1	3.249356	0.0714516	

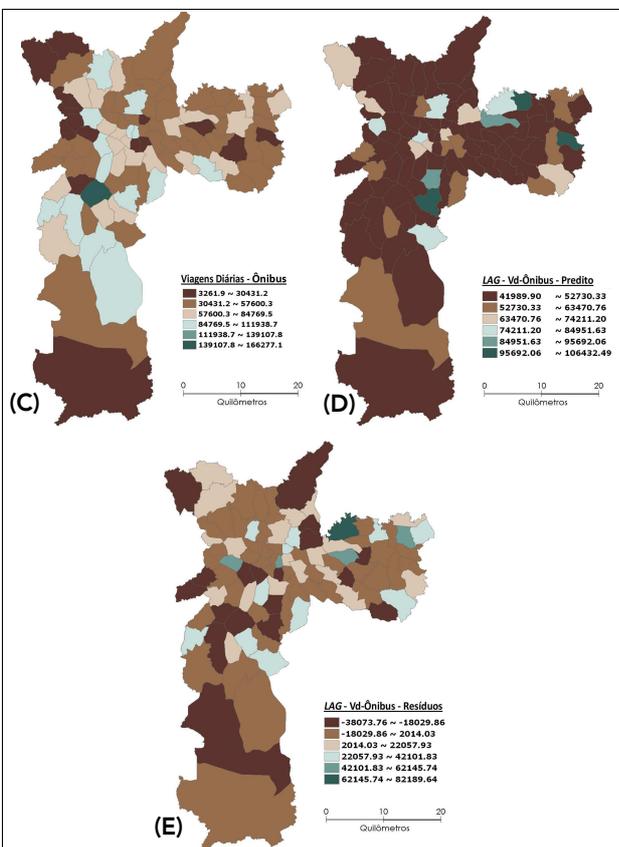
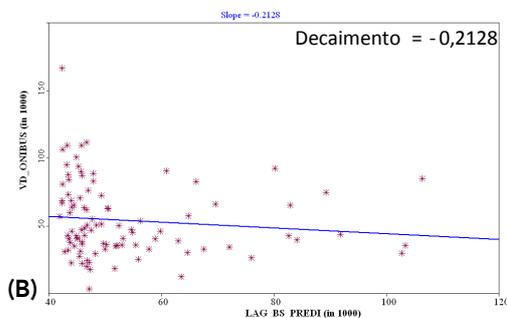


Figura 22

(A) Resultados da aplicação do *Spatial Error Model* no Modelo 2 (VD.ÔNIBUS x EMP.SERVIÇOS), (B) gráfico relacionando os valores originais e preditos e mapas cloropléticos mostrando: (C) número original de viagens diárias de ônibus, (D) valores de VD.ÔNIBUS preditos e (E) resíduos.

REGRESSION		SUMMARY OF OUTPUT: SPATIAL ERROR MODEL		
<b>- MAXIMUM LIKELIHOOD ESTIMATION</b>				
Data set	: Análise pol			
Spatial Weight	: pesos.GAL			
Dependent Variable	: VD_ONIBUS	Number of Observations	: 96	
Mean dependent var	: 53932.114583	Number of Variables	: 2	
S.D. dependent var	: 27103.347378	Degree of Freedom	: 94	
Lag coeff. (Lambda)	: 0.511811			
R-squared	: 0.368760 ↑	R-squared (BUSE)	: -	
Sq. Correlation	: -	Log likelihood	: -1097.198223	
Sigma-square	: 463703201.72	Akaike info criterion	: 2198.4 ↓	
S.E of regression	: 21533.8	Schwarz criterion	: 2203.525142	
Variable	Coefficient	Std. Error	z-value	Probability
CONSTANT	38626.97	0.000222128	1.738951e+008	0.0000000
E_SERVIC	1.113857	6.344718	0.1755566	0.8606423
LAMBDA	0.5118113	9.376678	0.05458344	0.9564702
REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	
Breusch-Pagan test	1	6.984748	0.0082207	
REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR SPATIAL DEPENDENCE				
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : pesos.GAL				
TEST	DF	VALUE	PROB	
Likelihood Ratio Test	1	12.01452	0.005279	

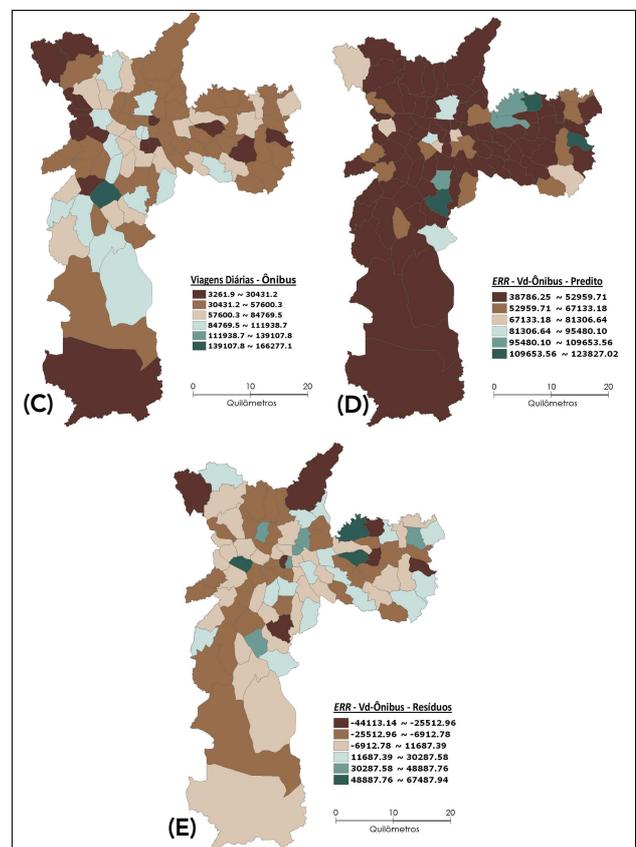
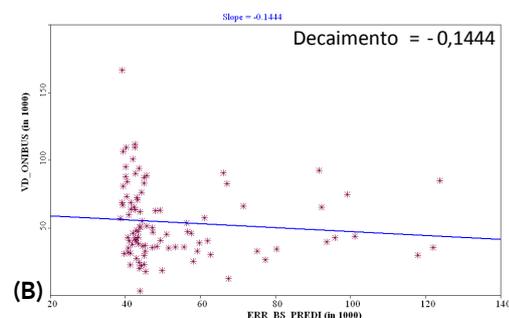


Figura 23

(A) Resultados da aplicação do *Spatial Lag Model* no Modelo 3 (VD.METRÔ x EMP. SERVIÇOS), (B) gráfico relacionando os valores originais e preditos e mapas coropléticos mostrando: (C) número original de viagens diárias de metrô, (D) valores de VD.METRÔ preditos e (E) resíduos.

REGRESSION - SUMMARY OF OUTPUT: SPATIAL LAG MODEL	
- MAXIMUM LIKELIHOOD ESTIMATION	
Data set	: Analise pol
Spatial Weight	: pesos.GAL
Dependent Variable	: VD_METRO
Mean dependent var	: 15968.4
S.D. dependent var	: 21715.7
Lag coeff. (Rho)	: 0.555242
Number of Observations	: 96
Number of Variables	: 3
Degrees of Freedom	: 93
R-squared	: 0.579006 ↑
Sq. Correlation	: -
Sigma-square	: 1.98529e+008
S.E of regression	: 14090
Log likelihood	: -1057.13
Akaike info criterion	: 2120.26 ↓
Schwarz criterion	: 2127.95

Variable	Coefficient	Std. Error	z-value	Probability
W_VD_METRO	0.5552424	11.26797	0.04927617	0.9606991
CONSTANT	230.5146	0.0006953831	331493	0.0000000
E_SERVIC	0.5187967	11.92645	0.04349968	0.9653031

REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	
Breusch-Pagan test	1	101.3821	0.0000000	

REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR SPATIAL DEPENDENCE				
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : pesos.GAL				
TEST	DF	VALUE	PROB	
Likelihood Ratio Test	1	28.2479	0.0000001	

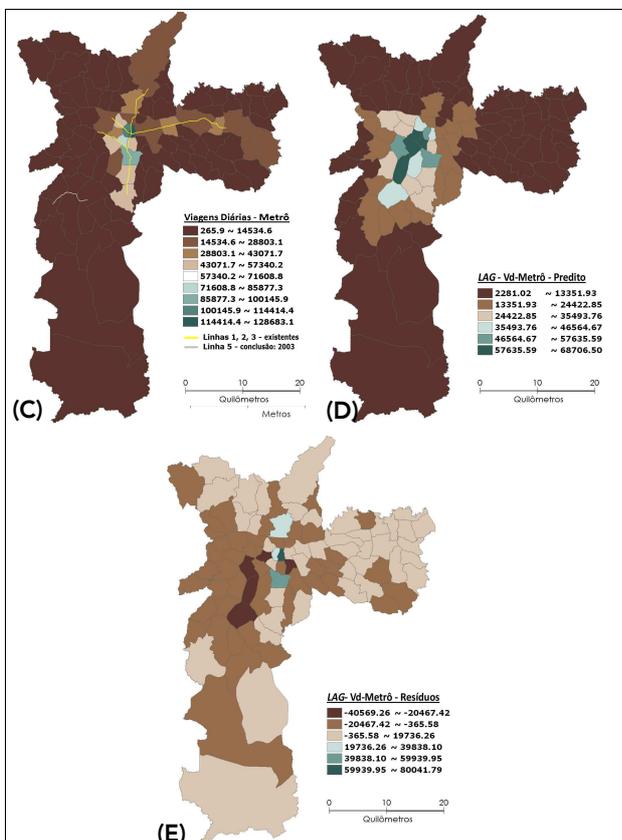
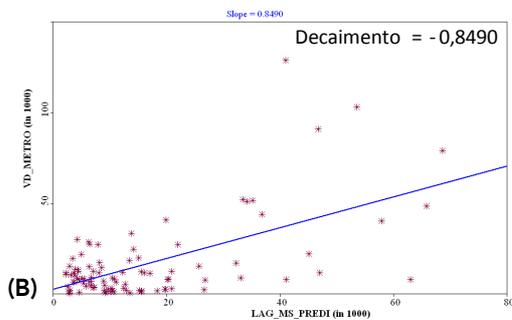


Figura 24

(A) Resultados da aplicação do *Spatial Error Model* no Modelo 3 (VD.METRÔ x EMP. SERVIÇOS), (B) gráfico relacionando os valores originais e preditos e mapas coropléticos mostrando: (C) número original de viagens diárias de metrô, (D) valores de VD.METRÔ preditos e (E) resíduos.

REGRESSION - SUMMARY OF OUTPUT: SPATIAL ERROR MODEL	
- MAXIMUM LIKELIHOOD ESTIMATION	
Data set	: Analise pol
Spatial Weight	: pesos.GAL
Dependent Variable	: VD_METRO
Mean dependent var	: 15968.416667
S.D. dependent var	: 21715.687214
Lag coeff. (Lambda)	: 0.697423
Number of Observations	: 96
Number of Variables	: 2
Degree of Freedom	: 94
R-squared	: 0.644912 ↑
Sq. Correlation	: -
Sigma-square	: 167449172.88
S.E of regression	: 12940.2
Log likelihood	: -1051.780262
Akaike info criterion	: 2107.56 ↓
Schwarz criterion	: 2112.689221

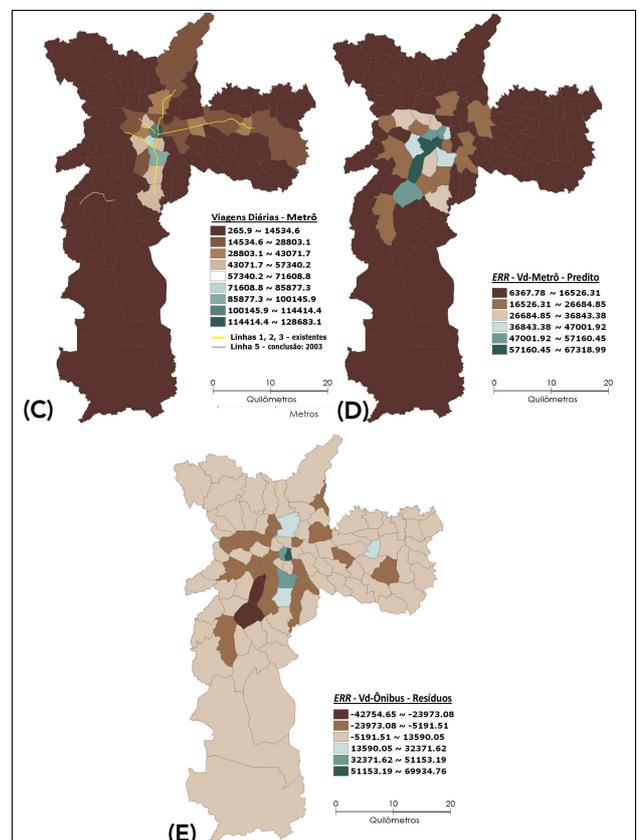
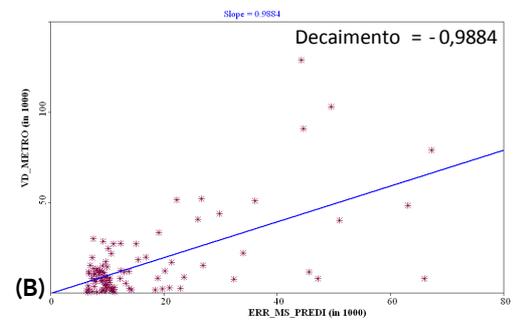
Variable	Coefficient	Std. Error	z-value	Probability
CONSTANT	6253.624	0.0002291026	2.729617e+007	0.0000000
E_SERVIC	0.798334	9.090273	0.08782289	0.9300173
LAMBDA	0.697423	12.77888	0.05457623	0.9564759

REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	
Breusch-Pagan test	1	121.8307	0.0000000	

REGRESSION DIAGNOSTICS - DIAGNOSTICS FOR SPATIAL DEPENDENCE				
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : pesos.GAL				
TEST	DF	VALUE	PROB	
Likelihood Ratio Test	1	38.94604	0.0000000	



não foram capazes de modelar a relação entre as variáveis resposta e explicativa, ou seja, as *viagens diárias realizadas pelo modo de transporte ônibus* não podem ser modeladas utilizando-se somente a variável independente *empregos no setor de serviços*. Os resíduos obtidos com a aplicação dos dois modelos espaciais também apresentaram-se insatisfatórios (Figuras 21E e 22E).

Com relação ao Modelo 3, a análise dos valores preditos pelos modelos espaciais, em relação aos valores observados (Figuras 23C, 23D, 24C e 24D) mostra que os modelos, uma vez mais, não foram capazes de modelar a relação entre as variáveis resposta e explicativa, ou seja, as *viagens diárias realizadas pelo modo de transporte metrô* não podem ser modeladas utilizando-se somente a variável independente *empregos no setor de serviços*. Entretanto, os resíduos obtidos com a aplicação dos dois modelos espaciais, apesar de insatisfatórios, apresentaram-se menos incoerentes, com uma variação menor de valores (Figuras 23E e 24E).

## 6. Análise dos Resíduos

Para se testar a dependência espacial dos resíduos, foi calculado o Índice de Moran e construído o *Moran Map* dos resíduos da regressão espacial.

O *Moran Map* exibe os locais onde ocorrem associações espaciais positivas (HH ou LL) ou negativas (HL ou LH). O *Lisa Map* mostra a significância desses agrupamentos (5%, 1%, 0,1% e 0,01%).

Para o Modelo 1 ( $VD.ÔNIBUS \times EMP.COMÉRCIO$ ), o índice de Moran dos resíduos apresentou os valores de 0,1338 (*Spatial Lag*) e 0,02 (*Spatial Error*), que indicam ausência de autocorrelação espacial (Figura 25).

O Modelo 2 ( $VD.ÔNIBUS \times EMP.SERVIÇOS$ ), os valores encontrados para o índice de Moran dos resíduos foram de 0,1009 (*Spatial Lag*) e

Figura 25

Moran Maps e Lisa Maps dos resíduos da regressão espacial pelo modelos *Spatial Lag* (A)(C) e *Error* (B);(D) no Modelo 1.

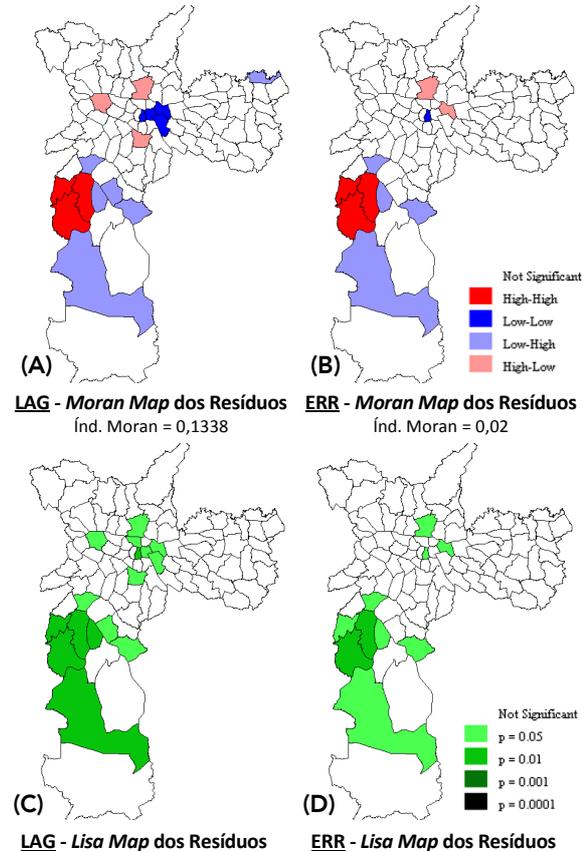
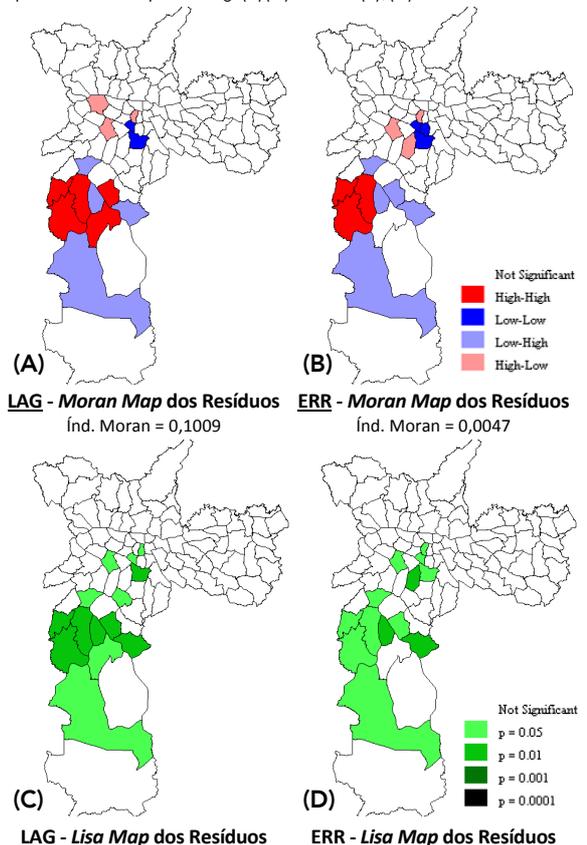


Figura 26

Moran Maps e Lisa Maps dos resíduos da regressão espacial pelo modelos *Spatial Lag* (A)(C) e *Error* (B);(D) no Modelo 2.

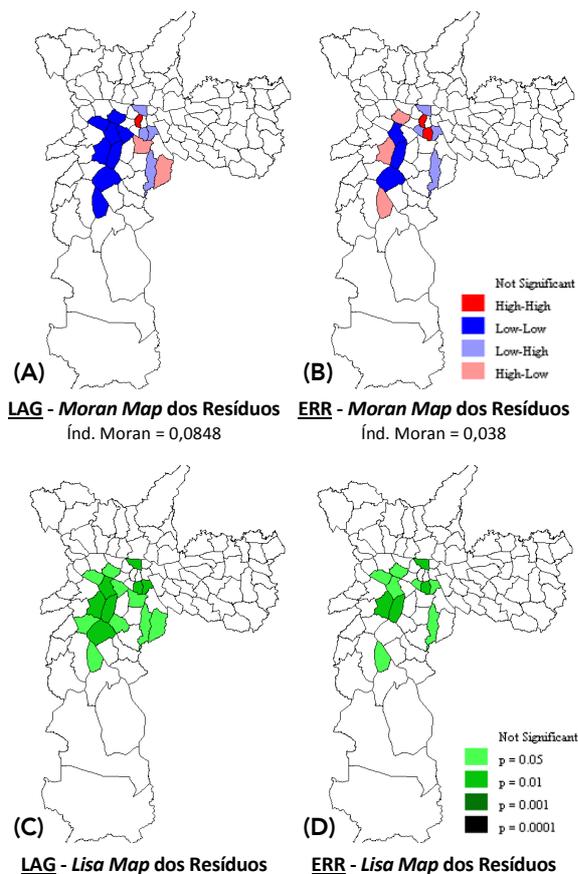


0,0047 (*Spatial Error*), que indicam ausência de autocorrelação espacial (Figura 26).

O Modelo 3 (VD.METRÔ × EMP.SERVIÇOS), os valores encontrados para o Índice de Moran dos resíduos foram de 0,0848 (*Spatial Lag*) e 0,038 (*Spatial Error*), que indicam ausência de autocorrelação espacial (Figura 27).

Figura 27

Moran Maps e Lisa Maps dos resíduos da regressão espacial pelo modelos *Spatial Lag* (A)(C) e *Error* (B)(D) no Modelo 3.



Ainda que um *Moran Map* possa exibir algum tipo de agrupamento, se esse agrupamento for pouco significativo, isto pode indicar ausência de dependência espacial dos resíduos. O Índice de Moran Global também deve entrar na discussão da existência de agrupamentos, pois, mesmo que o *Moran Map* e o *Lisa Map* indiquem alguma associação espacial, é necessário contrastar esse dado com o Índice de Moran Global, que pode reforçar a idéia de ausência de agrupamento.

## Conclusão

As análises realizadas no presente trabalho buscaram relacionar os dados da quantidade de viagens realizadas diariamente no município de São Paulo, utilizando os modos de transporte coletivos - ônibus, metrô e trem - com dados da oferta de empregos, especialmente no setor de serviços e comércio.

A aplicação dos modelos espaciais de regressão *Spatial Lag* e *Spatial Error*, embora tenham diminuído os valores de Akaike obtidos nas regressões lineares, não apresentaram significância, não melhorando, desta forma, os resultados obtidos pelo método de Regressão Linear.

Os resultados obtidos ao longo do desenvolvimento do trabalho mostraram que não é possível modelar o comportamento das viagens diárias por ônibus, metrô ou trem utilizando como variáveis independentes o número de empregos formais.

A principal dificuldade encontrada ao longo do desenvolvimento das análises refere-se à forma de agregação dos dados disponíveis: em Distritos oficiais do município de São Paulo. Através dos procedimentos realizados e das análises dos resultados obtidos, conclui-se que o comportamento e a distribuição das viagens (de um modo geral) não podem ser relacionados a nenhum tipo de dado sócio-econômico na escala utilizada no presente trabalho - a dos Distritos - para o caso do município de São Paulo.

Segundo Ramos (2002), a análise de dados territorializados é diretamente influenciada pela definição dos grupos de pesquisa da amostra ou espacialmente condicionados pela definição das fronteiras das subdivisões territoriais a que estes grupos de indivíduos estiverem associados. O estudo de qualquer variável territorializada deve, portanto, ser concebido em diferentes níveis de agregação de modo que resultados obtidos possam

corroborar para o verdadeiro entendimento do dado analisado.

Nesse sentido, estudos sobre demanda por infra-estrutura em deslocamento intra-urbano devem levar em consideração mais variáveis, de naturezas mais complexas e características de cada fração da cidade, área de estudo que deve ser composta de células tão pequenas quanto possível. Uma possível escala que pode melhor se adequar a este tipo de trabalho é a

das AEDs, do IBGE: Áreas de Expansão e Disseminação do Resultado da Amostra do Censo 2000.

Uma segunda recomendação refere-se às análises envolvendo o Metrô e os trens da CPTM: deve-se inserir, como forte restrição as próprias linhas férreas, ou seja, a não abrangência de todo o município, de forma equivalente ou adequada, pela malha metro-ferroviária.

## Referências

- ANSELIN, L., 1988. *Spatial Econometrics, Methods and Models*. Dordrecht, Kluwer Academic, 304p.
- ANSELIN, L., 1996. Simple diagnostic test for spatial dependence. *Regional Science and Urban Economics*, v. 26, n. 1, p. 77-104, Feb. 1996.
- ANSELIN, L., 2002. Under the hood: issues in the specification and interpretation of spatial regression models. *Agricultural Economics*. Disponível em: <http://agec221.uiuc.edu/users/anselin/papers/hood.pdf>
- BAILEY, T.C.; GATRELL, A. C., 1995. *Interactive Spatial Data Analysis*. 1ª ed. New York: Longman Scientific & Technical, v.1, 416 p.
- CÂMARA, G.; DAVIS, C.; MONTEIRO, A. M. V. (ed.), 2001. *Introdução à ciência da geoinformação*. São José dos Campos: INPE. Disponível em: <http://www.dpi.inpe.br/>
- CÂMARA, G.; MONTEIRO, A. M. V.; CARVALHO, M. S.; DRUCK, S., 2002. *Análise Espacial de dados Geográficos*, 2ª edição. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/>
- CHATERJEE, S; PRICE, B., 1977. *Regression Analysis by Example*. John Wiley & Sons, New York.
- FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M., 2000. *Quantitative Geography - Perspectives on Spatial Data Analysis*. London: Ed. Sage.
- IBGE, 2002. *Contagem da População 2002: documentação dos microdados da amostra*. Rio de Janeiro. Disponível em CD-ROM.
- SEMPLA – DEINFO. *INFOLOCAL: Sistema de Informação para o Planejamento Local*. Disponível em <http://www.infolocal.prefeitura.sp.gov.br>
- METRÔ - COMPANHIA DO METROPOLITANO DE SÃO PAULO. *Relatório da Pesquisa Origem-Destino do Metrô/SP de 1997*. Disponível em: [www.metro.sp.gov.br](http://www.metro.sp.gov.br)
- METRÔ - COMPANHIA DO METROPOLITANO DE SÃO PAULO. *Relatório da Aferição da Pesquisa Origem-Destino do Metrô/SP de 2002*. Disponível em: [www.metro.sp.gov.br](http://www.metro.sp.gov.br)
- NETER, J.; KUTNER, M. H.; NACHTSHEIM, C. J.; WASSERMAN, W., 1996. *Applied Linear Statistical Models*. 3ª ed. 1400 p.
- RAMOS, F. R., 2002. *Análise espacial de estruturas intra-urbanas: o caso de São Paulo*. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos.
- SMT, 2008. *São Paulo em Movimento*. <http://www9.prefeitura.sp.gov.br/spMovimento/sisnum/frotamunic.php>
- TASCHNER, S.; BOGUS, L. M., 2001. *São Paulo: Caleidoscópio urbano*. São Paulo em Perspectiva, jan./mar. 2001, v. 15, n. 1. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S010288392001000100005&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010288392001000100005&lng=en&nrm=iso)