Compatibilização de dados censitários para análises temporais com o auxílio de imagens Landsat

Flávia da Fonseca Feitosa
Antônio Miguel Vieira Monteiro
Gilberto Câmara

Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12201-970 - São José dos Campos - SP, Brasil
{flavia, miguel, gilberto}@dpi.inpe.br

Abstract. The changes in Census tracts boundaries over time represent a considerable restriction on temporal analysis based on this kind of data. This paper presents an alternative procedure to conciliate the geometry of census tracts established for different dates, as well as its associated data. For that, information extracted from satellite images is required. The proposed procedure was applied over the Census Data of São José dos Campos (SP) concerning the years 1991 and 2000.

Palavras-chave: Census Data, temporal analysis, remote sensing, dados censitários, análises temporais, sensoriamento remoto.

1. Introdução

São inúmeros os estudos que demandam dados provenientes de levantamentos populacionais cujas informações coletadas são relativas a indivíduos localizados em pontos específicos do espaço. Contudo, tais informações são de caráter confidencial e sua disponibilização exige cuidados, visto que a privacidade dos indivíduos deve ser preservada. Por esta razão, a solução normalmente adotada para a representação destes dados é sua agregação geográfica a unidades de área (Martin, 1996). Este procedimento parte da premissa de que as características populacionais apresentam um padrão discreto, que pode ser representado por polígonos fechados internamente homogêneos e com variações abruptas entre si. Sabe-se, porém, que normalmente o processo de delimitação destas unidades de área não objetiva a determinação de locais relativamente homogêneos, estando condicionados a critérios operacionais de levantamento ou políticos (Câmara et al., 2002).

No caso do Censo Demográfico Brasileiro, os registros individuais coletados são agregados em unidades de área conhecidas como setores censitários. De acordo com o IBGE, um setor censitário representa uma "unidade de controle cadastral formada por área contínua, situada em um único quadro urbano ou rural, com dimensão e número de domicílios que permitam o levantamento das informações por um único recenseador, segundo cronograma estabelecido" (IBGE, 2000b). Em outras palavras, a delimitação dos setores censitários, por definição, é guiada por critérios exclusivamente operacionais. Em casos como este, no qual o estabelecimento das unidades de área desconsidera a distribuição espacial da população, os resultados gerados sobre os dados são influenciados pela definição espacial das fronteiras destas unidades. Esta interferência é conhecida como *Problema das Unidades de Área Modificáveis* – MAUP (Openshaw, 1984).

Em análises temporais, as alterações nos resultados das análises provocadas pelo MAUP são ainda mais sérias, dado que as fronteiras dos setores censitários são geralmente modificadas a cada recenseamento. Estudos sob estas condições - sejam eles baseados em cálculos de índice, taxas, modelos de regressão etc. - tornam-se inviáveis, pois não é possível distinguir se as mudanças indicadas pelos resultados de fato ocorreram ou se representam meras distorções provocadas pelas alterações nas fronteiras dos setores censitários.

Uma opção existente para trabalhos desta natureza é o agrupamento de setores censitários compatíveis nos dois censos utilizando o conceito de áreas mínimas comparáveis (IPEA, 2004). Contudo, em geral este procedimento implica em uma agregação expressiva das informações associadas aos setores e conseqüente perda de detalhes fornecidos pelos dados originais.

Buscando amenizar este problema, este artigo apresenta um procedimento alternativo para a compatibilização da geometria dos setores censitários de datas distintas, bem como dos dados a eles associados. Diferentemente do critério utilizado para o estabelecimento de áreas mínimas comparáveis, que envolve apenas a agregação de áreas, o procedimento aplicado inclui também a desagregação de alguns setores. Esta desagregação é possível graças à integração de dados populacionais a dados do meio físico. Para tanto, imagens de satélite são utilizadas na identificação e quantificação das áreas ocupadas por usos urbanos nos anos dos levantamentos. Para ilustrar o procedimento, foram compatibilizados dados censitários dos anos de 1991 e 2000 relativos à área urbana do município de São José dos Campos (SP).

2. Compatibilização dos dados censitários de São José dos Campos

Conforme o exposto na seção anterior, cada setor censitário deve ser dimensionado de tal forma que permita o levantamento das informações por um único recenseador em um determinado tempo. Nas áreas urbanas, estes condicionantes resultam em setores com aproximadamente 250 domicílios. Em virtude deste critério de zoneamento, o aumento da população nas cidades vem acompanhado de um aumento no número de setores censitários, o que impõe dificuldades a estudos temporais.

No caso de São José dos Campos, cujo incremento da população no período 1991-2000 foi de 96.943 habitantes, a diferença entre o número de setores censitários dos dois últimos recenseamentos foi significativa. Em 1991, o município era subdividido em 356 setores censitários e, em 2000, passou a ser subdividido em 759 setores. A **Figura 1** mostra a localização do município de São José dos Campos e a área selecionada para a compatibilização dos setores de 1991 e 2000, que corresponde à área urbana e de expansão urbana da cidade (PMSJC, 2003). A área selecionada era subdividida em 347 setores no ano de 1991 (340 setores inteiros e 7 parciais) e passou a ser subdividida em 739 setores no ano de 2000.

Para a compatibilização das bases foram utilizados os seguintes dados:

- Censos Demográficos de São José dos Campos de 1991 e 2000 (IBGE, 1991, 2000).
- Tabelas de comparabilidade entre os setores censitários de 1991 e 1996 e entre os setores de 1996 e 2000 (IBGE, 1996, 2000c). Estas tabelas foram fornecidas pelo IBGE de São Paulo e descrevem a equivalência dos setores censitários em relação aos setores do levantamento anterior, assim como sua formação e o motivo das alterações ocorridas.
- Imagens Landsat-5/TM (INPE, 1990) e Landsat-7/TM (INPE, 2000) nas bandas 3 (0,63 μm - 0,69 μm), 4 (0,76 μm - 0,90 μm) e 5 (1,55 μm - 1,75 μm). As imagens correspondem às datas 27/04/90 e 09/01/00, respectivamente, e à órbita/ponto 219/76. A utilização de uma imagem do ano de 1990, ao invés de 1991, foi necessária em virtude da presença de nuvens nas imagens adquiridas durante o ano do recenseamento.
- Mosaico digital ortorretificado (PMSJC, 2003), referente ao ano de 2000 e na escala de 1:30.000.
- Dados complementares da cidade, mapas digitais do sistema viário e das quadras na escala de 1:2.000 (PMSJC, 2003).

Figura 1 – Localização de São José dos Campos (SP) e indicação da área selecionada para a compatibilização dos dados censitários.
Fonte: Adaptado de IBGE (2000) e INPE (2000).

2.1 Registro das imagens Landsat e delimitação das áreas de ocupação urbana

As imagens Landsat-5/TM (INPE, 1990) e Landsat-7/ETM+ (INPE, 2000) foram registradas tomando-se como referência o mosaico digital ortorretificado do ano 2000. Para o registro das imagens, foram utilizados polinômios de primeiro grau e seis pontos de controle bem distribuídos sobre a área de interesse.

Após as operações de registro, as áreas de ocupação urbana puderam ser extraídas das imagens. Para tanto, optou-se por uma classificação por interpretação visual sobre composições coloridas formadas pelas bandas 5,4 e 3.

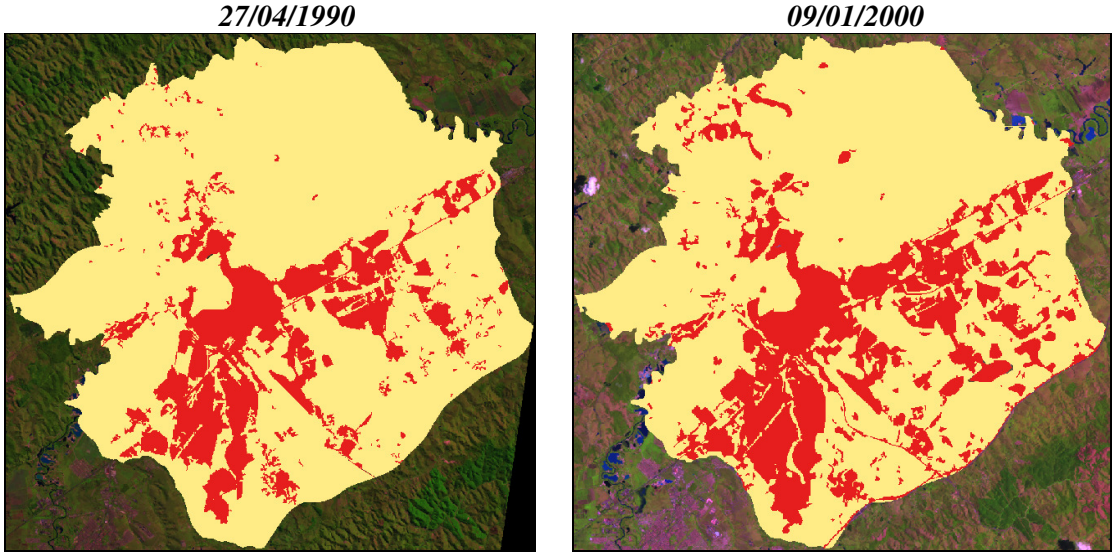

Figura 2 - Delimitação das manchas urbanas (em vermelho) sobre as imagens Landsat-5/TM (27/04/90) e Landsat-7/ETM+ (09/01/00).
Fonte: INPE (1990, 2000).

A classificação foi realizada com o auxílio dos mapas digitais do sistema viário e das quadras sobrepostos às imagens. Este procedimento permitiu uma delimitação mais precisa das áreas urbanas, dado que a diversidade dos materiais presentes nestas áreas por vezes dificulta sua identificação, mesmo quando realizada de maneira não-automática. A **Figura 2** indica, em vermelho, as áreas de ocupação urbana em cada uma das imagens. Os procedimentos de registro e classificação foram realizados no software SPRING (Câmara et al., 1996).

2.2 Compatibilização das bases geográficas dos setores censitários de 1991 e 2000

Após o registro e classificação sobre as imagens Landsat, uma prévia compatibilização das bases geográficas dos setores censitários de 1991 e 2000 foi realizada. Este procedimento resultou em uma nova base, constituída por setores provenientes da base de 1991 e de 2000. Nas áreas urbanas já consolidadas em 1991, a nova base foi constituída pelos setores desta data, ao passo que as demais áreas foram subdivididas pela geometria dos setores criados em 2000. A **Figura 3** ilustra o processo de elaboração desta base compatibilizada, indicando, em vermelho, os setores provenientes da base de 1991 e, em amarelo, os setores provenientes da base de 2000.

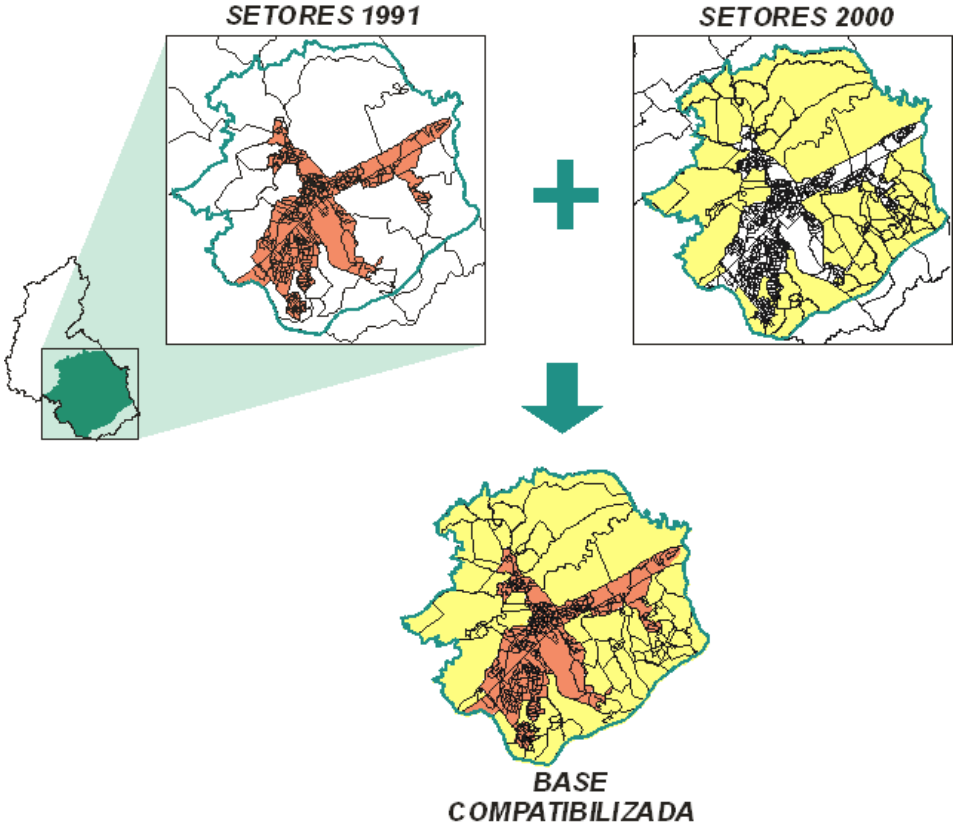

Figura 3 – Processo de elaboração da base geográfica compatibilizada.

Em virtude de erros inerentes às bases cartográficas originais, a sobreposição das mesmas gerou uma série de polígonos espúrios, ou seja, polígonos resultantes da intersecção de linhas que representam as mesmas feições, mas que têm representações digitais discrepantes (D'Alge, 2001). Nestes casos, é recomendável uma análise sobre os dados originais e a escolha da melhor representação como base para a digitalização (D'Alge, 2001). Comparando

com o mapa do sistema viário e as ortofotos da área, a geometria dos setores de 2000 foi considerada mais confiável do que a dos setores de 1991 e, por conseguinte, utilizada como referência na edição vetorial da base compatibilizada. Esta base, originalmente elaborada no software AutoCAD, foi importada para o software SPRING e a cada um de seus polígonos foi associado um código único.

2.3 Elaboração de tabelas de equivalência entre os polígonos da base compatibilizada e os setores originais

Após a compatibilização da base cartográfica, iniciou-se a etapa de elaboração de uma tabela de equivalência entre os polígonos desta nova base e os setores censitários de 1991 e 2000. Esta tabela foi elaborada com o auxílio das tabelas de comparabilidade dos setores 1991-1996 e 1996-2000 e das áreas urbanas existentes nos anos de 1990 e 2000, previamente delimitadas sobre as imagens Landsat (seção 2.1).

A tabela de equivalência elaborada apresenta os seguintes campos: (a) código dos polígonos da base compatibilizada (ID9100), (b) código do setor censitário de 1991 (ID1991), (c) peso do setor censitário de 1991 em relação ao polígono da base compatibilizada (PESO91), (d) código do setor censitário de 2000 (ID2000) e (e) peso do setor censitário de 2000 em relação ao polígono da base compatibilizada (PESO00).

A **Figura 4** mostra alguns registros da tabela de equivalência que ilustram polígonos da base compatibilizada cuja equivalência em relação aos setores de 1991 e 2000 é composta de diferentes maneiras. São estes:

ID9100	ID1991	PESO91	ID2000	PESO00
9100001	35499005000001	1	354990405000001	1
9100003	35499005000003	1	354990405000003	1
9100003	35499005000003	1	354990405000004	1
9100292	35499005000045	0,43	354990405000061	1

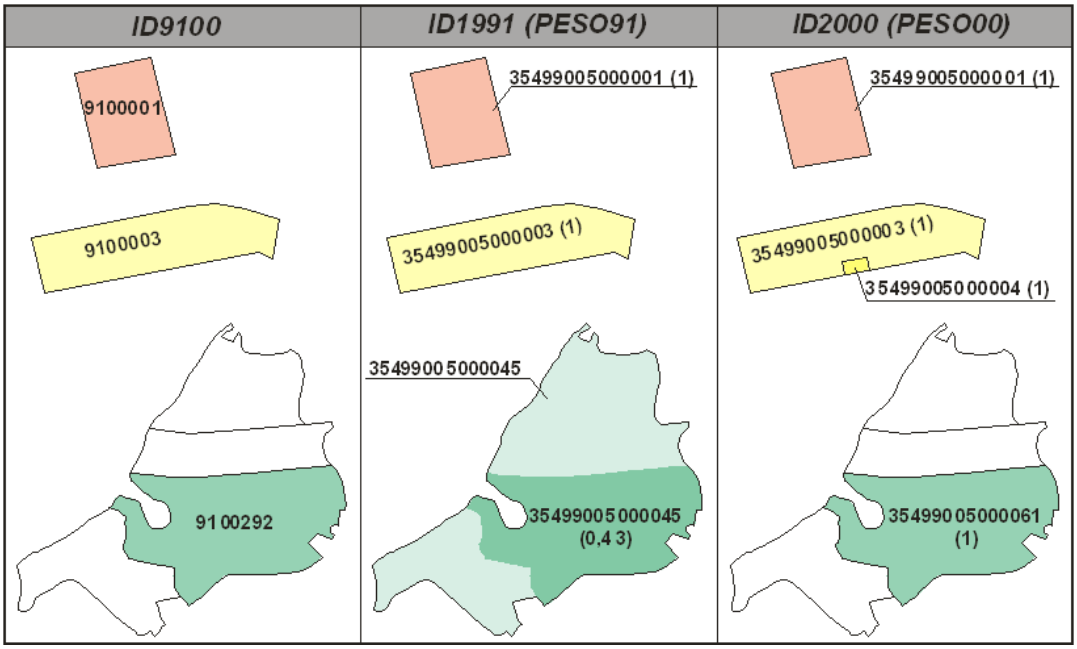

Figura 4 – Exemplos de registros da tabela de equivalência e geometria dos polígonos correspondentes.

(a) Polígonos cuja geometria equivale a de setores que não foram alterados no período 1991-2000. É o caso do polígono 9100001, conforme mostra a **Figura 4**.

(b) Polígonos cuja geometria equivale à de um setor de 1991 e de um ou mais setores de 2000, ou vice versa. É o caso do polígono 9100003. Em situações como esta, os dados dos setores de 2000 correspondentes ao polígono da nova base precisam ser agregados.

(c) Polígonos cuja geometria equivale a de um setor de 2000 e parte de um setor de 1991, ou vice versa. É o caso do polígono 9100292, para o qual os dados relativos ao setor de 1991 (35499005000045) precisam ser desagregados. Para esta desagregação é estabelecido um peso que equivale à proporção da área ocupada do setor 35499005000045 que está contida dentro dos limites do polígono 9100292. No caso, a área ocupada refere-se à realidade de 1990 e foi delimitada sobre a imagem Landsat-5/TM. A **Figura 5** mostra o setor 35499005000045 (1991) sobre a mancha urbana de 1990. Nesta data, a área ocupada dentro do polígono 9100292 corresponde a 43% da área ocupada do setor 35499005000045.

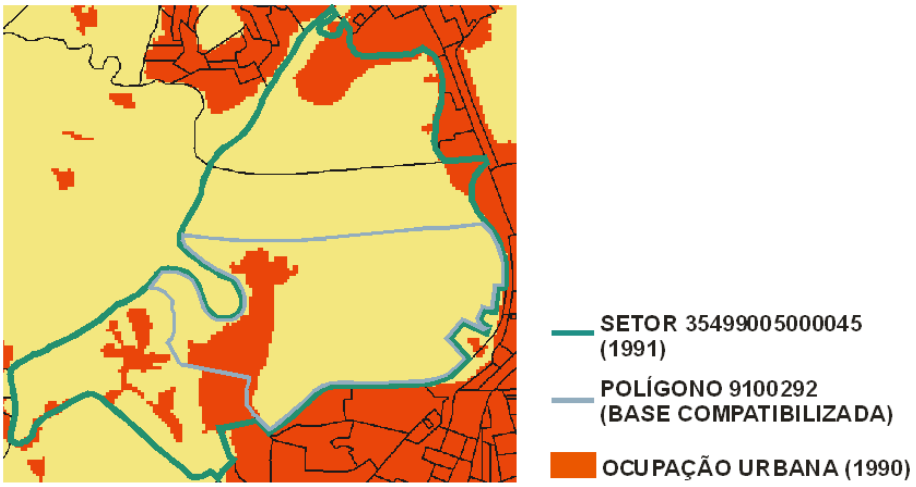

Figura 5 - Setor 35499005000045 (1991) sobre a mancha urbana de 1990.

(d) Existem ainda polígonos resultantes de situações mistas, que envolvem tanto a agregação quanto a desagregação dos setores de 1991 e 2000.

No processo de elaboração da tabela de equivalência, alguns polígonos da base geográfica compatibilizada (seção 2.2) foram revistos e por vezes agregados. A versão final da base compatibilizada apresenta 421 polígonos.

Uma vez finalizada, a tabela de equivalência foi subdividida em duas: a primeira contendo dos campos relativos à equivalência entre a base compatibilizada e os setores de 1991 (ID9100, ID1991 e PESO91) e a segunda contendo os campos relativos à equivalência entre a base compatibilizada e os setores de 2000 (ID9100, ID2000 e PESO00). Sobre cada uma destas tabelas, foi efetuado o seguinte: (a) exclusão dos registros repetidos, (b) junção com as tabelas contendo os dados censitários do ano ao qual os setores correspondem, (c) ponderação dos dados censitários utilizando o peso do setor e (d) arredondamento dos dados censitários ponderados.

Como resultado, foram obtidas tabelas contendo os códigos dos polígonos da base compatibilizada e os dados censitários correspondentes a cada um deles nos anos de 1991 e 2000. Estas tabelas foram associadas à base geográfica compatibilizada no SPRING, gerando

dois planos de informações cadastrais: um deles contendo a base compatibilizada associada aos dados censitários de 1991 e o outro contendo a mesma base compatibilizada, porém associada aos dados censitários de 2000.

3. Conclusões

Este artigo apresenta um procedimento para a compatibilização de dados censitários de datas distintas e agregados em setores com geometrias diferenciadas. Ao contrário da solução normalmente adotada nestes casos, a de áreas mínimas comparáveis, a alternativa proposta permite minimizar a perda de detalhes fornecidos pelos levantamentos originais. Para tanto, é necessária a integração de dados do meio físico, extraídos através de imagens de satélite, aos dados originais, de natureza sócio-demográfica.

O procedimento foi aplicado aos dados censitários da área urbana de São José dos Campos e mostrou-se de viável execução. Contudo, algumas limitações podem ser apontadas. A primeira delas diz respeito à desagregação dos dados ponderada pelas áreas ocupadas por usos urbanos em geral, embora os dados censitários sejam relativos somente às áreas providas de domicílios. Além disso, a desagregação é realizada apenas em função da área ocupada, desconsiderando a existência de densidades diferenciadas. Para minimizar estas limitações, a compatibilização das bases censitárias pode incorporar, quando disponíveis, dados provenientes de outras fontes, como mapas de uso do solo (identificação de áreas residenciais) ou cadastro de imóveis (identificação de domicílios).

Cabe salientar ainda, que a compatibilização entre os dados censitários não representa uma solução aos efeitos do MAUP, visto que este é um problema inerente a todos os dados agregados em áreas. Porém, o procedimento evita que inferências sejam realizadas sobre dados agregados em áreas de geometrias distintas, o que tende a provocar distorções inadmissíveis sobre os resultados obtidos.

Referências

Câmara, G.; Souza R.C.M.; Freitas, U. M.; Garrido, J. C. P. SPRING: Integrating remote sensing and GIS with object-oriented data modelling**. Computers and Graphics**, v.15, n.6, p. 13-22, 1996.

Câmara, G.; Monteiro, A. M.; Fucks, S. D.; Carvalho, M. S. **Análise Espacial e Geoprocessamento** [on-line], 2002. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/cap1-intro.pdf>. Acesso em: 18 ago. 2004.

D'Alge, J. C. L. **Cartografia para Geoprocessamento** [on-line], 2001. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/introd/cap6-cartografia.pdf>. Acesso em: 21 fev. 2004.

Instituto Brasileiro de Geografia e Estatística (IBGE). **Censo Demográfico 1991: São José dos Campos**. São Paulo**,** 1991.

_____. **Censo Demográfico 2000: São José dos Campos**. São Paulo**,** 2000.

_____. **Instruções para Revisão do Trabalho da Base Operacional Geográfica TR-23**. Rio de Janeiro: IBGE, 2000b. 33 p.

_____. **Tabela de Comparabilidade: Setores Censitários 1991-1996**. São Paulo**,** 1996.

_____. **Tabela de Comparabilidade: Setores Censitários 1996-2000**. São Paulo**,** 2000c.

Instituto de Pesquisa Econômica Aplicada (IPEA). [online]. **Dicionário de Conceitos**. Disponível em: <http://www.ipeadata.gov.br/ipeaweb.dll/Dicionario.htm>. Acesso em: 05 set. 2004.

Instituto Nacional de Pesquisas Espaciais (INPE). **Landsat-5/TM satellite image – WRS 219/76 – Acquisition date 27-04-90**. [CD-ROM]. São José dos Campos: INPE, 1990.

_____. **Landsat-7/ETM+ satellite image – WRS 219/76 – Acquisition date 09-01-00**. [CD-ROM]. São José dos Campos: INPE, 2000.

Martin, D. An assessment of surface and zonal models of population. **International Journal of Geographical Information Systems**, v. 10, p. 973-989, 1996.

Openshaw, S. Ecological fallacies and the analysis of areal census data. **Environment and Planning A**, v.16, p. 17-31, 1984.

Prefeitura Municipal de São José dos Campos (PMSJC). **Cidade Viva**. [CD-ROM], 2003.