

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

Modelos de Regressão para Estimação de Populações via Imagens de Satélite

Vanessa Loureiro Silva
Ilka Afonso Reis
Edna Afonso Reis

RELATÓRIO TÉCNICO
SÉRIE PESQUISA
RTP 03/2009

Modelos de Regressão para Estimação de Populações via Imagens de Satélite

Vanessa Loureiro Silva*

Ilka Afonso Reis**

Edna Afonso Reis**

* Graduanda em Estatística - UFMG

* Profa. Adjunta do Depto. de Estatística - UFMG

RESUMO

O tamanho da população humana de uma região é uma característica dinâmica. No Brasil, as contagens populacionais são realizadas em intervalos de 5 anos. Em regiões de acelerada ocupação imobiliária, por exemplo, o tamanho da população experimenta mudanças significativas em intervalos de tempo menores do que 5 anos. Desse modo, surge a necessidade da estimação de populações em intervalos menores, utilizando metodologia mais barata e rápida do que um recenseamento populacional. Diante do desenvolvimento da área de Sensoriamento Remoto, uma alternativa para estimação de populações humanas é utilizar imagens de satélite em modelos de regressão. A técnica estatística atualmente utilizada é baseada em um modelo de regressão linear gaussiano, com abordagem clássica de inferência. A variável dependente é a contagem (ou densidade) populacional e as variáveis explicativas são os valores médios de reflectância (proporção de energia refletida por um alvo) associados aos pixels da imagem nas várias bandas de um sensor. Esta modelagem pode ser feita utilizando como unidades de análise as unidades de coleta dos dados (como setores censitários) ou os pixels da imagem. Neste trabalho, propomos estimar o tamanho populacional dos setores censitários urbanos do município de Belo Horizonte, em 2000, via imagens de satélite e via informações da contagem populacional de 1996, através de modelos (por pixels e por setor) mais sofisticados que possam explorar de maneira mais abrangente a riqueza dos dados das imagens de satélite e dos censos populacionais.

Palavras-chave: Sensoriamento Remoto; Reflectâncias; Setores Censitários; Pixels.

Sumário

1	Introdução	3
2	Conceitos Básicos de Sensoriamento Remoto	4
3	Materiais e Métodos	6
3.1	O Conjunto de Dados	6
3.2	Modelos de Regressão	6
3.2.1	Abordagem Atual	6
3.2.2	Abordagens Propostas	8
4	Análise Estatística	10
4.1	Análise Exploratória	10
4.1.1	Banco de Dados dos Setores	10
4.1.2	Banco de Dados dos Pixels	12
4.2	Análise dos Modelos de Regressão	12
4.2.1	Banco de Dados dos Setores	12
4.2.2	Banco de Dados dos Pixels	22
5	Conclusões	23
6	Referências	25

1 Introdução

O tamanho da população humana de uma região é uma característica de grande relevância para o desenvolvimento de estudos em diversas áreas do conhecimento. Tal característica também apresenta um grande dinamismo, o que exige a atualização constante de seu registro.

No Brasil, os recenseamentos populacionais são feitos em intervalos de 5 anos, já que, a cada dez anos, o Instituto Brasileiro de Geografia e Estatística (IBGE) realiza um censo populacional e, a cada cinco anos, no intervalo entre dois censos, realiza uma contagem populacional. No entanto, em regiões de acelerada ocupação imobiliária, por exemplo, o tamanho da população experimenta mudanças significativas em intervalos de tempo menores do que cinco anos. Deste modo, surge a necessidade da estimação de populações em intervalos menores utilizando uma metodologia mais barata e rápida do que um recenseamento populacional. A estimação do tamanho da população nos anos em que não há censo e nem contagem populacional é usualmente feita através de técnicas de projeção demográfica.

Diante do desenvolvimento da área de Sensoriamento Remoto, uma alternativa para estimação de populações humanas é utilizar imagens de satélite em modelos de regressão. A técnica estatística atualmente utilizada é baseada em um modelo de regressão linear gaussiano, com abordagem clássica de inferência (Harvey, 2002a). A variável dependente é a contagem (ou densidade) populacional e as variáveis explicativas são os valores médios de reflectância (proporção de energia refletida por um alvo) associados aos pixels da imagem nas várias bandas de um sensor.

Essa modelagem pode ser feita em diversos níveis de agregação, utilizando como unidades de análise as unidades de coleta dos dados (como setores censitários) ou os pixels da imagem. Geralmente, o tamanho da unidade de coleta de dados é superior ao tamanho do pixel da imagem utilizada, ou seja, uma unidade de coleta contém vários pixels. Adotando esta abordagem, Harvey (2002a) utilizou imagens do sensor Thematic Mapper (TM), a bordo do satélite LANDSAT 5, para estimar a população dos CD's (Collect Districts) de duas regiões situadas na Austrália.

Harvey (2002b) utilizou os pixels da imagem como unidades de análise. Para contornar o problema da não existência de dados na população associados aos pixels, um procedimento denominado *regressão iterada* foi utilizado para distribuir a população do setor entre os pixels associados a ele. Comparou, então, o desempenho dos modelos de regressão assim estimados com aqueles estimados apenas ao nível dos CD's (Harvey, 2002a). Utilizando uma abordagem similar à de Harvey (2002b), Reis e Reis (2007) estimaram a população dos setores censitários urbanos do município de Belo Horizonte (MG), em 1996. As autoras conseguiram bons resultados no nível macro, superando o desempenho de modelos descritos na literatura. Porém, o desempenho no nível micro não pode ser considerado satisfatório. Portanto, a técnica tradicional pode ser melhorada.

Neste trabalho, propomos estimar o tamanho populacional dos setores censitários urbanos

do município de Belo Horizonte, em 2000, via imagens de satélite e via informações da contagem populacional de 1996, através de modelos (por pixels e por setor) mais sofisticados que possam explorar de maneira mais abrangente a riqueza dos dados das imagens de satélite e dos censos populacionais.

2 Conceitos Básicos de Sensoriamento Remoto

Sensoriamento remoto pode ser definido como o conjunto de técnicas que utilizam a radiação eletromagnética como meio de detectar e medir algumas características dos alvos de interesse. O Espectro Eletromagnético (EEM) é a ordenação contínua da radiação eletromagnética em função do comprimento de onda ou da frequência. A radiação eletromagnética contém, em ordem crescente de frequência de onda (ou decrescente do comprimento), ondas de rádio, microondas, radiação infravermelha, luz visível, radiação ultravioleta, raios-X e radiação Gama.

Nossa principal fonte de radiação é o Sol. Ao atingir um objeto (alvo), a radiação solar pode ser refletida, absorvida ou transmitida por este objeto. A proporção da energia refletida por um alvo é chamada reflectância. Cada alvo reflete diferentemente a energia em cada comprimento de onda do EEM. Desta forma, a reflectância de um alvo pode ajudar na caracterização do mesmo.

Uma imagem de satélite é formada por vários elementos menores denominados pixels, que representam uma porção do terreno imageado. Cada pixel contém a medida da radiação solar que foi refletida pela porção do terreno em determinada faixa de comprimento de onda (banda). Quanto mais clara a imagem naquela banda, maior é a reflectância nos comprimentos de onda daquela banda.

De estudos anteriores (Harvey(2002), Reis e Reis(2007)), observou-se que as reflectâncias nas várias bandas registradas pela imagem de satélite de um terreno podem auxiliar a estimar o tamanho populacional deste local. Os modelos de regressão a serem desenvolvidos a seguir utilizam esta idéia.

Tabela 1: Descrição dos bancos de dados finais.

Banco de Dados BH_2000_1 (Setores)	Banco de Dados BH_2000_2 (pixels)
Atributos :	Atributos:
1. Identificação dos setores	1. Identificação do pixel
2. Tipo de setor	2. Identificação do setor ao qual pertence o pixel
3. População total do setor em 2000	3. Tipo de setor ao qual pertence o pixel
4. População total do setor em 2000 morando em casas	4. Classificação do pixel em 2000: 0 - não urbano ; 1 - urbano ;
5. População total do setor em 2000 morando em apartamentos	5. Classificação do pixel em 1996: 0 - não urbano ; 1 - urbano ;
6. População estimada de 1996 (nos setores de 2000)	6. População total em 2000 do setor ao qual pertence o pixel
7. Densidade populacional em 1996 (considerando apenas área urbana)	7. População estimada de 1996 do setor ao qual pertence o pixel
8. Densidade populacional em 2000 (considerando apenas area urbana)	8. Densidade populacional do setor em 1996 (área urbana)
9. Reflectância média nas 6 bandas (6 colunas)	9. Densidade populacional do setor em 2000 (área urbana)
10. Média das razões entre bandas (15 colunas)	10. Reflectância de superfície nas 6 bandas (6 colunas)
11. Área total do setor em 1996 (estimada via pixels da imagens)	11. Área total do setor em 1996 (estimada via pixels da imagens)
12. Área total do setor em 2000 (estimada via pixels da imagens)	12. Área total do setor em 2000 (estimada via pixels da imagens)
13. Área urbana em 1996 (estimada via imagens)	13. Área urbana do setor em 1996 (estimada via imagens)
14. Área urbana em 2000 (estimada via imagens)	14. Área urbana do setor em 2000 (estimada via imagens)
15: Coordenada X do centróide do setor	15: Coordenada X do centróide do setor
16. Coordenada Y do centróide do setor	16. Coordenada Y do centróide do setor
	17. Colunas da matriz
	18. Linhas da matriz

3 Materiais e Métodos

3.1 O Conjunto de Dados

A construção dos bancos de dados (por setor e por pixels) foi feita em Ortiz *et al.* (2009). Para tal, foi necessária a obtenção de três tipos de dados: o cadastro urbano do município de Belo Horizonte com a divisão em setores censitários de 1996 e 2000, a contagem populacional de 1996 e o censo de 2000, por setor censitário, e a imagem das bandas 1 a 5 e 7 do sensor Thematic Mapper (TM) a bordo do satélite LANDSAT (de 22/09/2000). Como houve modificações nos setores censitários de 1996 para 2000, estimou-se a população de 1996 nos setores de 2000 por meio da espacialização da população de 1996 e posterior soma desta população por pixel dentro dos setores de 2000. As variáveis dos bancos de dados estão listadas na Tabela 1.

3.2 Modelos de Regressão

Nesta seção, são apresentados os modelos de regressão linear utilizados no contexto desse trabalho. Uma referência mais extensa sobre modelos de regressão lineares pode ser encontrada em Draper e Smith (1998).

3.2.1 Abordagem Atual

A técnica atualmente utilizada para estimação de populações é baseada em um modelo de regressão linear gaussiano, com abordagem clássica (Harvey(2002), Reis e Reis (2007)). A variável dependente é a contagem (ou densidade) populacional e as variáveis explicativas são as reflectâncias (ou reflectâncias médias) nas bandas. As unidades de análise são os setores censitários ou os pixels da imagem.

O modelo de regressão ao nível dos setores censitários

Nessa abordagem, as unidades geográficas básicas são os setores censitários. Como os setores têm áreas diferentes, é preferível que a variável dependente do modelo de regressão seja a densidade populacional, e não a contagem populacional.

Nos modelos usuais, a densidade populacional p_i de cada setor indexado por i , $i = 1, \dots, N$, é modelada por:

$$p_i = \beta_0 + \sum_{j=1}^k \beta_j r_{ji} + \epsilon_i, \quad (1)$$

onde os erros ϵ_i são assumidos independentes com distribuição Normal com média zero e variância constante e a variável explicativa r_{ji} é a média da reflectância dos pixels da imagem, na j -ésima banda do sensor ($j = 1, \dots, k$), associados ao setor i . Modelos mais complexos podem utilizar outros tipos de combinações das reflectâncias, como razão de bandas,

por exemplo, e também outras variáveis associadas (ex. porcentagem de área urbana) como variáveis explicativas.

Transformações na variável dependente também podem ser incorporadas ao modelo. Usualmente, a distribuição da densidade populacional dos setores é assimétrica, o que justifica o uso de uma transformação que leve à simetria e possibilite o ajuste do modelo Gaussiano. As transformações mais comuns para contornar a assimetria são a raiz quadrada e o logaritmo natural.

O modelo de regressão ao nível dos pixels

Nessa abordagem, as unidades geográficas básicas são os pixels. O modelo de regressão ao nível dos pixels tem a mesma fórmula do modelo apresentado em (1). Entretanto, r_{ij} é a reflectância na j -ésima banda do sensor ($j = 1, \dots, k$) em cada pixel indexado por i , $i = 1, \dots, M$, e p_i representa a população (ou densidade populacional) no i -ésimo pixel. Surge, então, uma limitação, visto que os registros populacionais disponíveis estão associados aos setores censitários e não aos pixels.

A solução para este problema foi proposta por Harvey (2002b) e denominada por ele de **regressão iterada**. A idéia da regressão iterada é redistribuir a população do setor em seus pixels através de sucessivas regressões. Inicialmente, a população do setor deve ser igualmente distribuída entre seus pixels, ou seja, a estimativa inicial de população em cada pixel em um dado setor é dada por:

$$p_i = P/n, \quad i = 1, 2, \dots, n,$$

onde P é a população do setor e n é o número de pixels do setor. Em seguida, ajusta-se o modelo de regressão em (1), obtendo-se os valores estimados \hat{p}_i para p_i . Deve-se, então, corrigir estas estimativas para que a população total do setor se mantenha no valor conhecido. O ajuste é dado pela seguinte expressão:

$$p_{i(ajust)} = \hat{p}_i + \bar{r}, \quad i = 1, 2, \dots, n, \quad (2)$$

onde \bar{r} é a média dos resíduos no setor, ou seja,

$$\bar{r} = \frac{\sum_{i=1}^n (p_i - \hat{p}_i)}{n}.$$

As estimativas ajustadas em uma iteração substituem as estimativas iniciais na iteração seguinte. A equação de regressão é novamente estimada, novos valores de \hat{p}_i são calculados e ajustados por (2), e assim por diante. As estimativas são ajustadas e assim por diante. O critério de parada pode ser definido em função de alguma medida de qualidade de ajuste, como o coeficiente de determinação (R^2). Assim, o procedimento deve parar quando o incremento no valor de R^2 passa a ser muito pequeno de uma iteração para outra. A última equação é então utilizada para gerar as estimativas de população.

Segundo Harvey (2002b), a multicolinearidade entre as reflectâncias nas bandas do TM pode causar problemas de convergência deste processo de regressão iterada. Outro problema lógico é a possibilidade de gerar estimativas negativas para a população dos pixels, já que o modelo de regressão linear utilizado não restringe os valores dos \hat{p}_i 's aos positivos. Na tentativa de solucionar este problema, o autor utilizou o seguinte procedimento: a cada iteração, as estimativas negativas eram transformadas em zero e ajustes nas estimativas dos outros pixels eram feitos de modo a manter constante o total populacional do CD.

Outro problema associado aos modelos de regressão ao nível dos pixels, como salientado em Harvey (2002b), é que os bancos de dados de pixels podem ser muito grandes, significando muito esforço computacional nas análises. Para solucionar este problema, Harvey sugere que se trabalhe com uma amostra dos pixels.

3.2.2 Abordagens Propostas

Apresentamos, a seguir, extensões do modelo de Harvey em (1). A primeira incorpora um componente autoregressivo. A segunda e a terceira, além da inclusão deste componente, utilizam, respectivamente, o Método de Mínimos Quadrados Generalizados e a Regressão Geograficamente Ponderada para estimação dos parâmetros.

Modelo via Mínimos Quadrados Ordinários com componente autorregressivo

O modelo em (1) não aproveita as informações de contagens populacionais de anos anteriores para a estimação da densidade populacional de determinado ano, as quais, em geral, são altamente correlacionadas. Assim, uma primeira modificação no modelo em (1) é modelar a densidade populacional (por pixels ou por setores) no tempo t da seguinte forma:

$$p_{(i,t)} = \eta p_{(i,t-1)} + \beta_0 + \sum_{j=1}^k \beta_j r_{ji} + \epsilon_i, \quad (3)$$

onde os erros ϵ_i são assumidos independentes com distribuição Normal com média zero e variância constante e a variável explicativa r_{ji} é a reflectância (ou reflectância média) dos pixels da imagem, na j -ésima banda do sensor ($j = 1, \dots, k$), associados ao pixel (ou setor) i e $p_{(i,t-1)}$ é a densidade populacional do pixel (ou setor) i no tempo $(t - 1)$. Na abordagem por pixels, utiliza-se a regressão iterada mencionada anteriormente para obter as estimativas populacionais ao nível dos pixels.

Modelo via Mínimos Quadrados Generalizados com componente autorregressivo

Aplicações envolvendo imagens de satélite são intrinsicamente de natureza espacial, dado que a energia atribuída a um pixel pode sofrer a influência de pixels vizinhos. Ao assumir erros independentes, o modelo em (3) pode estar violando uma suposição do método de estimação e omitindo informação espacial relevante para o seu bom ajuste. A análise espacial dos

resíduos do modelo dado em (3) deve ser feita, a fim de verificar a necessidade de se permitir que os erros sejam correlacionados entre si no espaço. Se verificada tal necessidade, propomos ajustar um modelo com a mesma expressão de (3), mas os erros ϵ_i são modelados como erros normais dependentes no espaço. Assim, neste modelo, estimado via Mínimos Quadrados Generalizados, propomos que a correlação entre os erros ϵ_i e ϵ_j seja dada pelo inverso da distância $d(i, j)$ entre os centróides do i -ésimo e do j -ésimo pixel (ou setor), ou seja,

$$Var(\epsilon) = \sigma^2 \mathbf{V} = \sigma^2 \begin{bmatrix} 1 & 1/d(1,2) & \dots & 1/d(1,n) \\ 1/d(2,1) & 1 & \dots & 1/d(2,n) \\ & & \dots & \\ 1/d(n,1) & 1/d(n,2) & \dots & 1 \end{bmatrix}, \quad (4)$$

com \mathbf{V} simétrica e não-singular.

Modelo com efeito espacial local

Dada a natureza espacial dos dados de imagens de satélite e a provável não-estacionariedade deste processo, na modelagem de tais dados, seria interessante que os coeficientes de regressão refletissem a heterogeneidade espacial. Uma alternativa neste caso é modelar a tendência espacial de forma contínua, com parâmetros variantes no espaço, realizando a chamada Regressão Geograficamente Ponderada (GWR, do inglês *Geographically Weighted Regression*). A GWR é uma técnica relativamente simples que estende a regressão tradicional. Baseia-se no fato de que o relacionamento entre a variável resposta e as variáveis explicativas pode ser diferente para cada uma das unidades de análise. Desta maneira, é como se cada i -ésima unidade de análise (pixel ou setor) tivesse seu modelo de regressão. Como não há dados suficientes para estimar tantos parâmetros, as estimativas dos coeficientes são feitas com base na informação dos vizinhos, dentro de uma largura de banda (*bandwidth*) pré-especificada. Assim, ficam todas dependentes umas das outras. Logo, a densidade populacional na i -ésima unidade de análise no tempo t seria modelada da seguinte forma:

$$p_{(i,t)} = \eta_i p_{(i,t-1)} + \beta_{i0} + \sum_{j=1}^k \beta_{ij} r_{ji} + \epsilon_i, \quad (5)$$

Deve-se salientar, entretanto, que esta técnica tem um poder mais explicativo do que preditivo. Mais detalhes sobre a GWR podem ser obtidos em Brunson *et al* (1996).

4 Análise Estatística

4.1 Análise Exploratória

4.1.1 Banco de Dados dos Setores

De acordo com os dados originais, no ano de 2000, Belo Horizonte era dividida em 2564 setores censitários.

Os setores especiais (presídios, asilos, orfanatos, etc) foram excluídos da análise sob a justificativa de que a contagem populacional nestes setores é relativamente fácil de ser obtida, já que geralmente existe um controle sobre o número de residentes nestes estabelecimentos. Alguns setores com valores sem sentido (estimativas negativas para população, etc) também foram excluídos.

Desta maneira, permaneceram 2528 setores, dos quais três eram considerados rurais. Por representarem um número pequeno, os três setores censitários rurais foram retirados do banco de dados, restando apenas os setores urbanos para serem analisados. Assim, para ajustar os modelos de regressão, os bancos de dados finais ficaram com 2525 setores, dos quais 309 (12,23%) são aglomerados sub-normais (favelas).

A Figura 1 a seguir apresenta a distribuição de frequências das densidades populacionais (em habitantes/10.000 m^2) dos setores urbanos nos anos de 1996 e 2000. Nota-se, para ambos os anos, uma acentuada assimetria à direita, ou seja, uma concentração nos valores mais baixos das densidades populacionais.

As estatísticas descritivas para as densidades populacionais dos setores censitários urbanos de Belo Horizonte em 1996 e em 2000 estão na Tabela 2 a seguir.

Tabela 2: Estatísticas descritivas para as densidades populacionais dos setores urbanos (em hab/10.000 m^2), nos anos 1996 e 2000.

Ano	Mínimo	Q1	Mediana	Média	Q3	Máximo	CV
1996	1,0	70,0	107,0	123,2	148,0	898,0	0,7480
2000	2,0	84,0	118,0	151,4	173,0	1800,0	0,8463

Da Tabela 2 e também da Figura 1, observa-se que as densidades populacionais dos setores urbanos de Belo Horizonte aumentaram razoavelmente de 1996 para 2000.

As estatísticas descritivas para as reflectâncias médias dos setores urbanos nas seis bandas são apresentadas na Tabela 3 a seguir:

Da Tabela 3, nota-se que as bandas na região do visível (bandas 1, 2 e 3) apresentam as menores reflectâncias médias, ao passo que as bandas na região do infravermelho (bandas

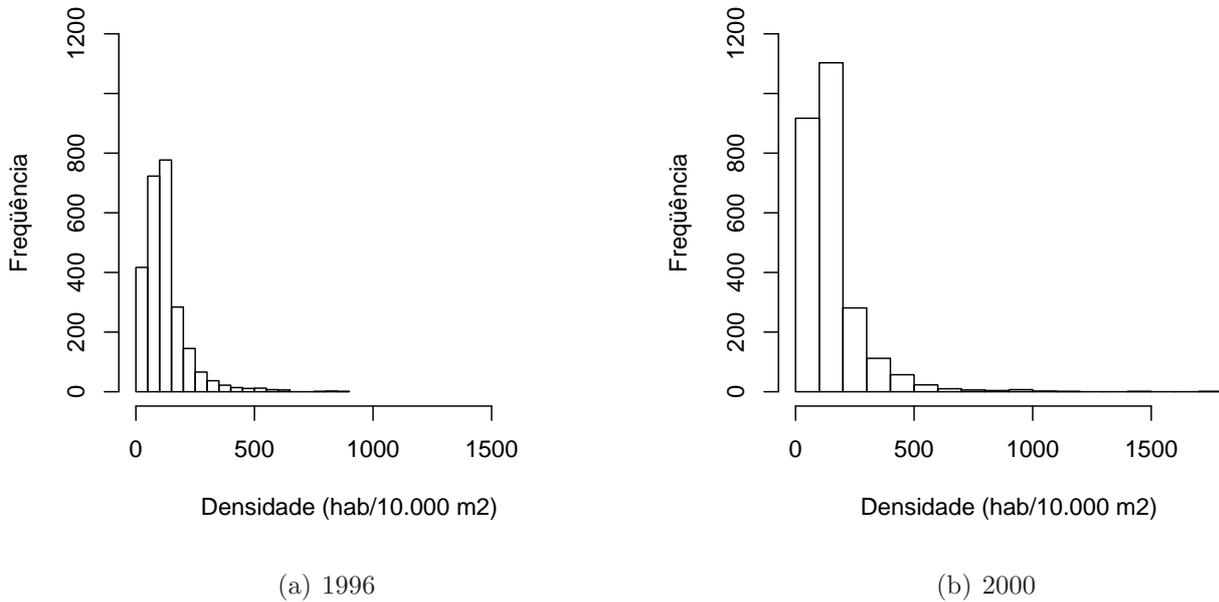


Figura 1: Distribuição de frequências das densidades populacionais dos setores urbanos (em hab/10.000 m^2), nos anos 1996 e 2000.

4, 5 e 7) apresentam as maiores. Em relação à variabilidade relativa, todas as bandas se assemelham.

A matriz com os coeficientes de correlação linear de Pearson entre as variáveis mais relevantes neste estudo é dada a seguir:

	Dens00	Dens96	B1	B2	B3	B4	B5	B7
Dens00	1,00							
Dens96	0,62	1,00						
B1	0,08	0,04	1,00					
B2	-0,01	-0,04	0,96	1,00				
B3	-0,06	-0,08	0,92	0,98	1,00			
B4	-0,22	-0,23	0,66	0,81	0,84	1,00		
B5	-0,17	-0,16	0,76	0,87	0,92	0,92	1,00	
B7	-0,09	-0,07	0,83	0,89	0,93	0,79	0,95	1,00

Em relação às correlações entre a variável densidade populacional no ano de 2000 e as principais variáveis candidatas a predictoras nos modelos de regressão, a variável densidade populacional no ano de 1996 é a que possui a maior correlação linear com a resposta (0,62), o que seria um indício forte de sua importância para explicar/prever esta variável. Já dentre as candidatas a predictoras, observa-se uma alta correlação entre vários pares de reflectâncias médias. Em especial, nota-se que as reflectâncias médias das bandas 2 e 3 apresentam correlações altas entre si e com todas as demais bandas.

Tabela 3: Estatísticas descritivas para as reflectâncias médias dos setores urbanos nas seis bandas da imagem do LANDSAT-TM.

Banda*	Mínimo	Q1	Mediana	Média	Q3	Máximo	CV
1 (450 a 520)	2,6810	12,9000	14,7300	14,5000	16,2400	25,0000	0,1970
2 (520 a 600)	5,2320	19,9100	22,4500	22,0600	24,7100	36,8200	0,1903
3 (630 a 690)	5,9690	21,3800	24,3000	23,8400	26,7700	41,3700	0,1989
4 (760 a 900)	14,2000	38,7400	44,2100	43,4700	48,6500	68,2100	0,1864
5 (1550 a 1750)	13,4000	42,4700	48,4400	47,0300	52,8500	71,5700	0,1908
7 (2080 a 2350)	11,5800	39,1100	44,5700	43,0800	48,5600	66,2000	0,1879

*Faixa espectral (em nm).

4.1.2 Banco de Dados dos Pixels

Conforme já mencionado anteriormente, os bancos de dados finais, com a retirada dos setores especiais e urbanos, ficaram com 2525 setores, o que corresponde a um total de 290780 pixels.

As estatísticas descritivas para as reflectâncias dos pixels urbanos nas seis bandas são apresentadas na Tabela 4.

Tabela 4: Estatísticas descritivas para as reflectâncias dos pixels urbanos nas seis bandas da imagem do LANDSAT-TM (na escala de 0 a 255).

Banda*	Mínimo	Q1	Mediana	Média	Q3	Máximo	CV
1 (450 a 520)	1	12	14	14,3938	17	76	0,3439
2 (520 a 600)	0	18	22	22,5209	27	254	0,2987
3 (630 a 690)	1	19	24	24,5626	28	146	0,3220
4 (760 a 900)	0	41	46	47,6196	53	212	0,2246
5 (1550 a 1750)	0	43	50	50,3770	57	254	0,2526
7 (2080 a 2350)	0	37	44	44,2495	51	228	0,2665

*Faixa espectral (em nm).

Da Tabela 4, nota-se que as bandas na região do visível (bandas 1, 2 e 3) apresentam as reflectâncias com valores mais baixos e com maior heterogeneidade, enquanto as bandas na região do infravermelho (bandas 4, 5 e 7) apresentam as reflectâncias com valores mais altos e mais homogêneos.

A matriz com os coeficientes de correlação linear de Pearson entre as reflectâncias nas seis bandas são dadas a seguir:

	B1	B2	B3	B4	B5	B7
B1	1,00					
B2	0,91	1,00				
B3	0,84	0,95	1,00			
B4	0,29	0,48	0,52	1,00		
B5	0,43	0,56	0,63	0,69	1,00	
B7	0,61	0,69	0,74	0,45	0,85	1,00

As correlações lineares os pares das bandas 1, 2 e 3 são bastante elevadas, assim como a correlação entre as bandas 5 e 7.

4.2 Análise dos Modelos de Regressão

4.2.1 Banco de Dados dos Setores

Algumas medidas de avaliação comuns a todos os modelos ajustados, apresentadas nas tabelas das páginas seguintes, são explicadas a seguir:

- **Erro Relativo Mediano (ERM):**

O erro relativo para cada setor foi calculado pela diferença absoluta entre o valor observado da população e o valor estimado pelo modelo, como mostra a equação (6)

$$\text{Erro relativo} = \frac{|\text{observado-estimado}|}{\text{observado}}. \quad (6)$$

O erro relativo mediano é a mediana dos valores absolutos dos erros relativos e representa o erro no nível micro (setores).

- **Erro Relativo Total (ERT):**

O erro relativo total foi calculado como

$$\text{Erro relativo total} = \frac{\text{soma das estimativas de todos os setores}}{\text{soma dos observados para todos os setores}} - 1 \quad (7)$$

para representar a variação do total estimado em relação ao total observado, sendo chamado erro no nível macro (área urbana). Valores negativos para o erro relativo total indicam super-estimação e valores positivos indicam sub-estimação.

- R^2 “ajustado” (R_a^2):

Corresponde ao coeficiente de determinação da regressão “ajustado”, que leva em conta o número de variáveis explicativas em relação ao número de observações.

- R^2 “back” (R_b^2):

Corresponde ao quadrado do coeficiente de correlação entre as estimativas de população para os setores e os valores populacionais reais destes setores.

- **Análise dos sinais dos coeficientes das bandas 4, 5 e 7:**

Estas bandas estão localizadas na região do infravermelho, sendo as bandas 4 e 5 bastante utilizadas na detecção de vegetação. Por esta razão, seria esperado um sinal negativo para os coeficientes destas bandas (4 e 5): quanto maior a reflectância nestas bandas, maior a presença de vegetação e menor a presença da população. Com relação à banda 7, pixels com valores altos de reflectância indicariam a presença de superfícies construídas (Harvey 2002b), sendo esperado um sinal positivo para o coeficiente desta banda.

A Tabela 5 apresenta os resultados para diversos modelos ajustados via Mínimos Quadrados Ordinários.

Inicialmente, ajustou-se o modelo usual, ou seja, tendo apenas as reflectâncias médias como explicativas. Da Tabela 5, observa-se que tal modelo (1.1) tem um desempenho muito ruim. A inclusão da densidade populacional de 1996, modelo (1.3), melhorou o modelo (1.1). No entanto, observamos que os sinais das bandas 5 e 7 são contrários ao esperado, tanto no modelo (1.1) como no modelo (1.3). Lembrando que a distribuição da variável resposta tem uma assimetria acentuada (Figura 1(b)), a transformação logarítmica nas densidades populacionais de 1996 e 2000 foi adotada (modelo (1.4)), determinando sinais coerentes dos coeficientes de todas as bandas, a melhoria da estimação no nível micro e da análise dos resíduos (passam a ter uma distribuição mais “simétrica”) mas a piora da estimação no nível macro. Este último resultado, no entanto, não é preocupante, visto que este trabalho não tem o objetivo de estimar a população total da cidade. O modelo com a transformação logarítmica foi então refinado com a inclusão das variáveis Indicador de favela e Percentual de área urbana, resultando em uma melhoria pequena em seu desempenho (modelo (1.5)).

Um problema ocorrido no modelo (1.5) e nos anteriormente ajustados é a multicolinearidade, como se observa na Tabela 6 a seguir. De acordo com o critério sugerido por Draper e Smith (1998), os modelos que não apresentam sérios problemas de inflação da variância seriam aqueles cujos Fatores de Inflação da Variância (VIF's, na sigla em inglês) são menores do que 10. Da Tabela 6, nota-se que as bandas 2 e 3 são as que possuem os VIF's mais elevados (bem maiores do que 10) em todos os modelos. Como já visto anteriormente, as reflectâncias médias destas bandas são altamente correlacionadas com as das demais bandas. Desta maneira, optou-se por retirar tais bandas do conjunto de variáveis explicativas.

Seguindo Harvey (2002a), utilizou-se ainda o método de seleção automática *stepwise*, colocando como variáveis candidatas a preditoras, além das já apresentadas (Reflectâncias médias - exceto bandas 2 e 3 já retiradas, Logaritmo da Densidade de 1996, Indicador de favela e Percentual de área urbana), todos os pares de razões entre bandas. O último modelo apresentado na Tabela 5 - modelo (1.6) - é o resultante, com todas as variáveis significativas ao nível de significância de 5%. Da Tabela 6, observa-se que tal modelo não tem problemas de multicolinearidade.

Tabela 5: Medidas de avaliação para alguns dos modelos ajustados via Mínimos Quadrados Ordinários ao banco de dados de setores urbanos.

Modelo	R_a^2	R_b^2	ERT(%)	ERM(%)	Sinal R4	Sinal R5	Sinal R7	VIFs altos?
1.1 - Resposta: Densidade 2000 - Explicativas: 6 bandas	0,1497	0,1517	-8,88E-08	36,6094	Negativo	Positivo	Negativo	Sim
1.2 - Resposta: Log. Densidade 2000 - Explicativas: 6 bandas	0,1938	0,1155	21,8844	32,9203	Negativo	Negativo	Positivo	Sim
1.3 - Resposta: Densidade 2000 - Explicativas: 6 bandas, Densidade 96	0,4292	0,4308	-1,11E-07	26,5153	Negativo	Positivo	Negativo	Sim
1.4 - Resposta: Log. Densidade 2000 - Explicativas: 6 bandas, Log. Densidade 96	0,5107	0,3937	9,5430	23,0105	Negativo	Negativo	Positivo	Sim
1.5 - Resposta: Log. Densidade 2000 - Explicativas: 6 bandas, Log. Densidade 96 Indicador de favela, Área urbana (%)	0,5693	0,3935	7,8735	21,1161	Negativo	Negativo	Positivo	Sim
1.6 - Resposta: Log. da Densidade 2000 - Explicativas: Bandas 1e 4, Log. Densidade 96 Indicador de favela, Área urbana (%), Razões 13, 15, 23, 24, 25 e 57	0,5868	0,4042	7,5614	21,5493	Negativo	***	***	Não

Tabela 6: Fatores de inflação da variância (VIF's) dos modelos apresentados na Tabela 5.

Variável/Modelo	1.1 e 1.2	1.3	1.4	1.5	1.6
Reflectância B1	44,19	44,20	44,20	45,08	2,89
Reflectância B2	179,90	180,99	181,08	181,15	***
Reflectância B3	79,75	80,80	82,30	82,99	***
Reflectância B4	21,06	21,20	21,07	22,20	5,83
Reflectância B5	66,95	67,00	68,40	70,55	***
Reflectância B7	35,39	35,53	37,83	40,96	***
Densidade 1996	***	1,15	***	***	***
Log Densidade 1996	***	***	1,25	1,33	1,28
Indicador de Favela	***	***	***	1,15	1,12
Percentual de área urbana	***	***	***	1,98	1,92
Razão 13	***	***	***	***	1,28
Razão 15	***	***	***	***	1,56
Razão 24	***	***	***	***	1,18
Razão 25	***	***	***	***	1,85
Razão 57	***	***	***	***	3,04

Em relação à análise residual, foi calculado o índice I de Moran para os resíduos deste modelo, visto que tal índice representa uma medida de autocorrelação espacial apropriada para avaliar o quanto uma observação em uma área depende de valores de áreas vizinhas. Foram definidas duas listas de vizinhança, considerando-se vizinhos os setores cuja distância entre os centróides seria até 1000 m ou 2000 m. Os valores observados para o índice I de Moran segundo as duas vizinhanças foram, respectivamente, 0,1210 e 0,0399. Testes de permutação pelo método Monte Carlo, com 999 permutações, foram realizados para testar a significância dos Índices I de Moran observados. Portanto, foram testadas as seguintes hipóteses H_0 : não há autocorrelação espacial ($I = 0$) versus H_1 : Há autocorrelação espacial positiva ($I > 0$). Os valores p obtidos em ambos os casos foi de 0,001, o que determina que o modelo (1.6) não conseguiu captar toda a informação dos dados, ainda presentes nos resíduos.

Dessa maneira, em seguida, foram ajustados modelos via Mínimos Quadrados Generalizados, permitindo que os erros sejam correlacionados entre si no espaço, segundo a matriz dada em (4). As medidas de desempenho de alguns modelos ajustados estão na Tabela 7. Nota-se que nenhum modelo conseguiu desempenho muito diferente dos modelos apresentados na Tabela 5. A única medida que melhorou bastante em relação aos modelos via MQO foi o valor de R^2 “ajustado”. Os teste de Monte Carlo baseados no Índices I de Moran realizados com os resíduos de todos os modelos, considerando as duas listas de vizinhança, ainda apresentam valores p de 0,001, indicando que ainda há informação espacial nos resíduos.

Na tentativa de obter um modelo melhor, a Regressão Geograficamente Ponderada foi então aplicada aos dados. As medidas de desempenho de alguns modelos ajustados também estão apresentadas na Tabela 7 a seguir. Como foi ajustada uma regressão para cada setor, o valor do coeficiente R^2 “ajustado” neste caso representa a mediana do vetor de coeficientes

R^2 “ajustado” de cada regressão. Utilizando as mesmas variáveis do melhor modelo obtido via MQO (modelo 1.6), o modelo (3.3) mostrou o melhor desempenho dentre todos os modelos, com erros menores tanto ao nível micro como macro. No entanto, a análise espacial dos resíduos deste modelo, feita via um teste específico proposto por Leung et al. (2000) para testar autocorrelação espacial entre resíduos da GWR, ainda indica que os resíduos contém informações relevantes, já que os valores p dos testes, foram, para as duas listas de vizinhança, iguais a $2,73e-11$ e 0.0189 .

Finalmente, a última tentativa foi separar o banco de dados segundo as variáveis Indicador de Favela (Favela e Não-Favela) e Densidade de 1996 (separação de acordo com a sua mediana) e, utilizando as mesmas variáveis do melhor modelo obtido via MQO (modelo 1.6), ajustar novos modelos com os bancos desmembrados. Primeiramente, foram ajustados os modelos via MQO e foram retiradas as variáveis não significativas. Os modelos via MQG e GWR foram ajustados com as mesmas variáveis dos modelos via MQO. Os resultados estão nas Tabelas 8, 9 e 10 adiante.

Das Tabelas 8, 9 e 10, observa-se uma melhoria dos modelos para setores normais (não-favelas), e, principalmente, para os setores mais densos (Densidade de 1996 > mediana) em relação aos modelos com o banco todo. O contrário ocorre com os modelos para setores que são favelas e menos densos. Em relação à análise residual, como se observa na Tabela 11 a seguir, os valores p para os testes de autocorrelação espacial dos resíduos se elevaram um pouco para os modelos dos setores mais densos (modelos (4.4), (5.4) e (6.4)), sendo inclusive, maiores do que 0,05 para os modelos ajustados via MQO e MQG, o que representa uma consequência da melhoria dos modelos para setores mais densos.

Tabela 7: Medidas de avaliação para alguns dos modelos ajustados via MQG e GWR ao banco de dados de setores urbanos.

Modelo	R_a^2	R_b^2	ERT(%)	ERM(%)	Sinal R4	Sinal R5	Sinal R7
MQG							
2.1 - Resposta: Densidade 2000 - Explicativas: 6 bandas, Densidade 96, Indicador de Favela, Área Urbana (%)	0,7224	0,4725	0,0635	22,8969	Negativo	Positivo	Negativo
2.2 - Resposta: Log. Densidade 2000 - Explicativas: 6 bandas, Log. Densidade 96, Indicador de Favela, Área Urbana (%)	0,9827	0,3922	7,9151	21,2167	Negativo	Negativo	Positivo
2.3 - Resposta: Log. da Densidade 2000 - Explicativas: Bandas 1e 4, Log. Densidade 96, Indicador de Favela, Área Urbana (%), Razões 13, 15, 23, 24, 25 e 57	0,9834	0,4032	7,6000	21,7034	Negativo	***	***
GWR							
3.1 - Resposta: Densidade 2000 - Explicativas: 6 bandas, Densidade 96, Indicador de Favela, Área Urbana (%)	0,5149	0,6134	0,3527	21,7653	***	***	***
3.2 - Resposta: Log. Densidade 2000 - Explicativas: 6 bandas, Log. Densidade 96, Indicador de Favela, Área Urbana (%)	0,6418	0,5989	4,6016	17,5042	***	***	***
3.3 - Resposta: Log. Densidade 2000 - Explicativas: Bandas 1e 4, Log. Densidade 96, Indicador de Favela, Área Urbana (%), Razões 13, 15, 23, 24, 25 e 57	0,6699	0,5474	5,1900	17,4428	***	***	***

Tabela 8: Medidas de avaliação para alguns dos modelos ajustados via Mínimos Quadrados Ordinários ao banco de dados de setores urbanos - Separação pelas variáveis Indicador de favela e Densidade de 1996.

Separação	Modelo	R_a^2	R_b^2	ERT (%)	ERM (%)	Sinal R4	VIF's altos?
4.1	Favela	0,4185	0,3984	-8,2767	25,4535	Negativo	Não
	- Resposta: Log. Densidade 2000						
	- Explicativas: Bandas 1e 4, Log. Densidade 96, Área Urbana (%), Razões 25 e 57						
4.2	Não-favela	0,5945	0,4182	-9,8531	19,4223	Negativo	Não
	- Resposta: Log. Densidade 2000						
	- Explicativas: Bandas 1e 4, Log. Densidade 96, Área Urbana (%), Razões 13, 15, 24, 25 e 57						
4.3	Densidade de 96 \leq mediana	0,4831	0,3106	-12,5081	23,6720	Negativo	Não
	- Resposta: Log. Densidade 2000						
	- Explicativas: Bandas 1e 4, Log. Densidade 96, Indicador de Favela, Área Urbana (%), Razões 13, 23, 24 e 57						
4.4	Densidade de 96 $>$ mediana	0,6101	0,4934	-5,5236	12,6844	Negativo	Não
	- Resposta: Log. Densidade 2000						
	- Explicativas: Bandas 1e 4, Log. Densidade 96, Indicador de Favela, Área Urbana (%), Razões 13, 15, 23, 24, 25 e 57						

Tabela 9: Medidas de avaliação para alguns dos modelos ajustados via Mínimos Quadrados Generalizados ao banco de dados de setores urbanos - Separação pelas variáveis Indicador de favela e Densidade de 1996

Separação	Modelo	R_c^2	R_b^2	ERT(%)	ERM(%)	Sinal R4
5.1 Favela	- Resposta: Log. Densidade 2000 - Explicativas: Bandas 1e 4, Log. Densidade 96, Área Urbana (%), Razões 25 e 57	0,9830	0,3982	9,0305	25,4618	Negativo
5.2 Não-favela	- Resposta: Log. Densidade 2000 - Explicativas: Bandas 1e 4, Log. Densidade 96, Área Urbana (%), Razões 13, 15, 24, 25 e 57	0,9864	0,4186	6,4053	19,7080	Negativo
5.3 Densidade de 96 \leq mediana	- Resposta: Log. Densidade 2000 - Explicativas: Bandas 1e 4, Logaritmo da Densidade 96, Indicador de Favela, Área Urbana (%), Razões 13, 23, 24 e 57	0,9844	0,3101	9,4845	23,6457	Negativo
5.4 Densidade de 96 $>$ mediana	- Resposta: Log. Densidade 2000 - Explicativas: Bandas 1e 4, Log. Densidade 96, Indicador de Favela, Área Urbana (%), Razões 13, 15, 23, 24, 25 e 57	0,9933	0,4936	2,9596	12,7016	Negativo

Tabela 10: Medidas de avaliação para alguns dos modelos ajustados via Regressão Geograficamente Ponderada ao banco de dados de setores urbanos - Separação pelas variáveis Indicador de favela e Densidade de 1996

Separação	Modelo	R_a^2	R_b^2	ERT(%)	ERM(%)
6.1 Favela	Resposta: Log. Densidade 2000 Explicativas: Bandas 1e 4, Log. Densidade 96, Área Urbana (%), Razões 25 e 57	0,5465	0,4958	7,8239	22,0931
6.2 Não-favela	Resposta: Log. Densidade 2000 Explicativas: Bandas 1e 4, Log. Densidade 96, Área Urbana (%), Razões 13, 15, 24, 25 e 57	0,6723	0,5590	4,3924	15,7862
6.3 Densidade de 96 \leq mediana	Resposta: Log. Densidade 2000 Explicativas: Bandas 1e 4, Log.Densidade 96, Indicador de Favela, Área Urbana (%), Razões 13, 23, 24 e 57	0,6147	0,4474	5,6625	19,9088
6.4 Densidade de 96 $>$ mediana	Resposta: Log. Densidade 2000 Explicativas: Bandas 1e 4, Log. Densidade 96, Indicador de Favela, Área Urbana (%), Razões 13, 15, 23, 24, 25 e 57	0,6058	0,5020	2,6754	12,5533

Tabela 11: Valores p para os testes de autocorrelação espacial dos resíduos dos modelos das Tabelas 8, 9 e 10.

Modelo	Vizinhança 1 (1000 m)	Vizinhança 2 (2000 m)
4.1	0,001	0,001
4.2	0,001	0,001
4.3	0,001	0,001
4.4	0,053	0,060
5.1	0,001	0,001
5.2	0,001	0,001
5.3	0,001	0,001
5.4	0,071	0,053
6.1	0,001	0,000
6.2	0,000	0,001
6.3	0,003	0,043
6.4	0,031	0,031

Embora a análise residual dos modelos ajustados nesta seção ao nível dos setores censitários indique que os resíduos ainda contêm informações não captadas pelos modelos, em termos das medidas de desempenho, pode-se dizer que a abordagem por setor neste trabalho determinou resultados melhores do que os encontrados em (Reis e Reis, 2007). Na abordagem por setores, o melhor modelo de Reis e Reis (2007) obteve, ao nível micro, um erro relativo mediano de 27,8% (contra 17,4% do modelo (3.3)) e, ao nível macro, um erro relativo total de 8,6% (contra 5,2% do modelo (3.3)).

4.2.2 Banco de Dados dos Pixels

Como colocado em Harvey (2002b), os bancos de dados de pixels podem ser muito grandes, significando muito esforço computacional nas análises. Para solucionar este problema, Harvey sugere que se trabalhe com uma amostra dos pixels. Assim como em (Reis e Reis, 2007), neste trabalho, foram amostrados 25% dos pixels de cada setor censitário. Depois de amostrados os pixels de cada setor, uma fração de 25% da população do setor foi igualmente distribuída entre os pixels selecionados. Esta população inicial foi então redistribuída através da regressão iterada, como explicado na Seção 3.3.1. As medidas de desempenho de alguns modelos ajustados ao banco dos pixels via MQO estão na Tabela 12.

Da Tabela 12, observa-se que, nos modelos ao nível dos pixels, as reflectâncias não são variáveis explicativas relevantes. Tais modelos têm um desempenho bem inferior aos correspondentes ao nível dos setores, dados na Tabela 5, ao contrário do ocorrido em (Reis e Reis, 2007). Na abordagem por pixels, o melhor modelo de Reis e Reis (2007) obteve, ao nível micro, um erro relativo mediano de 30,4% (contra 35,1% do último modelo da Tabela 12) e, ao nível macro, um erro relativo total de 0,47% (contra 20,9% do último modelo da Tabela 12). Desta maneira, não se tentou ajustar os modelos mais sofisticados - Mínimos

Quadrados Generalizados e Regressão Geograficamente Ponderada - aos bancos de dados dos pixels.

Tabela 12: Medidas de avaliação para alguns dos modelos ajustados via Mínimos Quadrados Ordinários ao banco de dados dos pixels urbanos

Modelo	R^2 iterada*	R_b^2	ERT (%)	ERM (%)	Sinal R4	Sinal R5	Sinal R7	VIFs altos?
- Resposta: Dens. 2000 - Explicativas: 6 bandas	0,0214	0,0226	-98,4454	98,7468	-	-	+	Sim
- Resposta: Dens. 2000 - Explicativa: Dens. 1996	0,6580	0,3351	21,5641	35,7480	**	**	**	Não
- Resposta: Den.2000 - Explicativas: 6 bandas, Dens. 1996	0,6582	0,3351	21,8011	35,9698	-	+	-	Sim
- Resposta: Dens. 2000 - Explicativas: 6 bandas, Dens. 1996, Indicador de Favela, Área urbana (%)	0,6590	0,3351	20,8940	35,1070	-	+	-	Sim

* Valor do R^2 "ajustado" correspondente à última iteração.

5 Conclusões

Nesse trabalho, foram propostos alguns modelos para estimar o tamanho populacional dos setores censitários urbanos do município de Belo Horizonte, em 2000, via imagens de satélite e via informações da contagem populacional de 1996. Os modelos de regressão foram ajustados considerando duas unidades de análise - setores censitários e pixels (urbanos somente), como em Reis e Reis (2007).

Em Reis e Reis (2007), foi adotado um modelo simples, via Mínimos Quadrados Ordinários, com uma combinação linear das reflectâncias nas imagens bandas 1 a 5 e 7 do LANDSAT-TM, somente para setores urbanos, para explicar a Densidade Populacional do município de Belo Horizonte em 1996. Outras variáveis - Indicador de aglomerados (favelas) e Porcentagem de área urbana - foram incorporadas ao modelo inicial na tentativa de melhorá-lo. Na abordagem por setores, o melhor modelo de Reis e Reis (2007) não conseguiu bons resultados nem ao nível micro nem ao nível macro. Já na abordagem por pixels, o melhor modelo de Reis e Reis (2007) conseguiu bons resultados apenas no nível macro.

Nesse trabalho, os resultados foram contrários aos ocorridos em Reis e Reis (2007). A abordagem por setor mostrou-se melhor do que a abordagem por pixels. No entanto, o desempenho do melhor modelo, (3.3), ainda não pode ser considerado satisfatório (erros relativos ainda não estão tão baixos).

O fato de a abordagem por setores ter sido melhor do que em Reis e Reis (2007) pode ser justificado pelo acréscimo de uma variável explicativa extremamente relevante - a Densidade populacional por setor de 1996. Ao contrário do que aconteceu com a abordagem dos pixels (Reis e Reis, 2007), as densidades eram diferentes para cada unidade de análise (setor). Além disto, o fato de ter sido utilizada uma medida-resumo, a média, para representar as reflectâncias dos setores suavizou o efeito da maior heterogeneidade da reflectâncias dos pixels de um mesmo setor e o modelo ajustou-se melhor. Finalmente, outro fato é que a média das reflectâncias dos pixels de cada setor foi feita excluindo-se os pixels classificados como não urbanos, assim como a densidade foi calculada levando-se em conta a área classificada como urbana do setor. Esses fatores podem ter contribuído para melhorar a abordagem dos setores, a ponto de ela superar a abordagem por pixel. Isto representa uma vantagem, visto que a montagem do banco de dados por setor e a própria regressão são muito mais fáceis.

Já o fato de a abordagem por pixels ter sido pior do que em Reis e Reis (2007) pode ser justificado pela baixa explicação das reflectâncias por pixels. Outra explicação pode ser que a correção aplicada nas imagens de agora tenha gerado imagens com mais contraste (mais variabilidade entre as reflectâncias dos pixels de um mesmo setor). Além disto, a Densidade de 1996 por pixel era a mesma para todos os pixels do setor. Como todos os pixels foram associados a um mesmo valor de densidade em 2000, mesmo com a re-distribuição feita regressão iterada, o modelo não conseguiu gerar coeficientes que fossem condizentes com o que esperávamos deles (sinais corretos). Uma maneira de corrigir esta possível discrepância entre as imagens seria corrigir a imagem de 2000 em relação à imagem de 1996 e não independentemente.

Como trabalhos futuros, destacam-se a incorporação da informação espacial via modelos espaciais bayesianos, o que pode resultar na obtenção de modelos cujos resíduos apresentem autocorrelação espacial reduzida ou nula.

6 Referências

1. Brunsdon, C. ; Fotheringham, A. S. ; Charlton, M. E. (1996) Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4), 281-298.
2. Draper, N. R. ; Smith, H. *Applied Regression Analysis*, 3a. edição. John Wiley and Sons, EUA, 706 p, 1998.
3. Harvey, J. T. Estimating census district populations from satellite imagery: some approaches and limitations. *International Journal of Remote Sensing*, vol. 23, n. 10, p. 2071-2095, 2002a.
4. Leung, Y. ; Mei, C-L. ; Zhang, W-X. (2000) Testing for spatial autocorrelation among the residuals of the geographically weighted regression. *Environment and Planning A*, 32, 871-890.
5. _____. Population estimation models based on Individuals TM Pixels. *Photogrammetric Engineering and Remote Sensing*, vol. 68, n. 11, p. 1181-1192, 2002b.
6. Ortiz, R..J.F. Reis, I.A., Assunção, R.M. (2009) Construção de Banco de Dados para Estimção da População de Belo Horizonte via Imagens de Satélite. *Anais da XVIII Semana de Iniciação Científica da Universidade Federal de Minas Gerais, Belo Horizonte, MG*.
7. Reis, I. A. ; Reis, E. A. (2007). *Estimção da População dos Setores Censitários de Belo Horizonte Usando Imagens de Satélite, Relatório Técnico de Pesquisa (RTP-01/2007)*, Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte.