



Análise e previsão demográfica utilizando matrizes de crescimento e distribuição populacional intermunicipal

Jeronimo Oliveira Muniz*

Entre os desafios para a área de projeções demográficas estão a volatilidade do componente migratório – fundamental para a projeção de pequenas áreas –, a compatibilização entre projeções de pequenas e grandes áreas de forma consistente e a mensuração e inclusão da incerteza em cenários futuros de crescimento demográfico. O presente artigo lida com estes três desafios ao desenvolver e testar um novo método probabilístico de previsão populacional para os municípios paulistas. As vantagens do método proposto são: utilizar como insumo apenas as informações sobre o município de residência anterior e as distribuições populacionais dos últimos dois Censos; gerar intervalos de confiança para as populações projetadas; explicitar o papel dos fluxos migratórios na dinâmica de crescimento intermunicipal; e facilitar a elaboração de cenários contrafactuais e análises de sensibilidade usando matrizes de crescimento multirregional. Descrevem-se os padrões e tendências das trocas migratórias municipais utilizando ferramentas de análise e visualização espacial e identificam-se as áreas nas quais a migração é responsável por parte considerável da dinâmica demográfica observada. Mais de 95% das 572 populações municipais paulistas projetadas tiveram boa precisão e localizaram-se dentro dos intervalos de confiança previstos. Utilizaram-se os dados dos Censos Demográficos de 1980, 1991 e 2000.

Palavras-chave: Projeções. Migração. Município. Crescimento natural. Análise espacial.

* Universidade Federal de Minas Gerais (UFMG), Belo Horizonte-MG, Brasil (jeronimo@fafich.ufmg.br; <https://orcid.org/0000-0002-5697-9516>).

Introdução

Uma das principais demandas por demógrafos advém da necessidade de se conhecer o número de pessoas que existirão em determinado lugar e tempo futuros. Saber onde, quando e quantos indivíduos existirão é essencial para conhecer o tamanho do mercado consumidor e planejar a oferta de serviços de saúde, educação, transporte, fornecimento de água e luz, coleta de lixo e infraestrutura pública e privada. Avanços científicos no instrumental técnico e teórico para projetar (*project*) e prever (*forecast*) a população nos níveis global, nacional, regional e local têm contribuído para tornar as estimativas mais precisas e menos subjetivas (STOTO, 1983; PFLAUMER, 1988; LEE, 1992; LEE; TULJAPURKAR, 1994; LUTZ et al., 1999; ALHO; SPENCER, 2005; KEYFITZ; CASWELL, 2005). Três desafios, entretanto, ainda permeiam o debate na área de projeções. O primeiro está na dificuldade em se projetar a população de pequenas e pequeníssimas áreas. Isso deve-se à volatilidade do componente migratório e à dificuldade em modelá-lo e prevê-lo. O segundo desafio está atrelado à consistência e à integração de projeções feitas de maneira separada, mas que representam parte de um todo. Esse é o caso, por exemplo, de projeções populacionais de pequenas áreas que, quando agregadas, fornecem estimativas distintas daquelas realizadas para a área maior da qual fazem parte. Em teoria, esse problema é resolvido por tabelas multiestado, mas na prática a aplicação do método é proibitiva quando há muitos “estados” envolvidos na projeção. Por fim, o terceiro desafio para realizar projeções relaciona-se à incerteza futura e à falta de consenso metodológico para lidar com a mesma.

No Brasil, em particular, estes três desafios têm sido pouco investigados.¹ A principal meta desta pesquisa é contribuir para o debate na área de previsões populacionais por meio do desenvolvimento e teste de uma metodologia alternativa para se prever o crescimento demográfico dos municípios paulistas. O método sugerido tem a virtude de integrar consistentemente as projeções de áreas menores e maiores e, ao mesmo tempo, lidar com a incerteza inerente à dinâmica demográfica futura.

Sobre projeções e estimativas demográficas no Brasil

No Brasil, o método de projeção dominante é o das componentes, que trata separadamente a influência da fecundidade, mortalidade e migração nas taxas de crescimento demográfico. Esse método, desenvolvido por Whelpton (1928) e formalizado por Leslie

¹ Waldvogel (1996), Fgoli (1999), Fgoli et al. (2003), Jannuzzi (2007), Brito, Cavenaghi e Jannuzzi (2010) e Santos e Barbieri (2015) são algumas exceções.

(1945, 1948), baseia-se no pressuposto de que a desagregação do crescimento demográfico em componentes conduz à maior precisão das estimativas populacionais. Tal metodologia, utilizada pelo IBGE, pelo Cedeplar e pela Fundação Seade desde 1970, consiste em observar o comportamento das taxas de fecundidade, mortalidade e migração e, a partir delas, gerar cenários futuros de crescimento para a área-alvo a ser projetada. O método possui dois méritos: ser consistente e flexível para lidar com hipóteses alternativas de crescimento e com desagregações por idade, sexo, cor, ou quaisquer outros cortes populacionais de interesse; e ser adaptável para levar em conta extrapolações e equações estruturais para prever o comportamento das componentes. Essa última prática, entretanto, é pouco usual no Brasil. Por outro lado, o método das componentes demográficas é inapropriado para a estimação populacional de pequenas áreas (por exemplo, municípios), já que neste nível de análise as estimativas de fecundidade e mortalidade podem se tornar temporalmente instáveis e de difícil mensuração (FREIRE; ASSUNÇÃO, 1998; CAVENAGHI et al., 2004). Por isso, a solução adotada por centros e institutos oficiais de pesquisa (tais como o IBGE) consiste em projetar a população para áreas maiores (unidades da federação, por exemplo) e posteriormente efetuar uma distribuição proporcional ou interpolativa para estimar a população das áreas menores (DUCHESNE, 1987; IBGE, 2013, 2016; SMITH et al., 2001; UNITED NATIONS, 1956). No caso da Fundação Seade (s/d, p. 1), “o modelo de projeção considerado adota uma hierarquia que parte da projeção para o total do Estado e se desagrega em regiões administrativas e municípios”, sem levar em conta as especificidades das trocas migratórias intermunicipais que, em alguns casos, correspondem a uma parcela significativa do crescimento demográfico. Além disso, o crescimento da subpopulação fica determinado pelo da população total (SZWARCOWALD; CASTILHO, 1989).

Os modelos tradicionalmente usados para as pequenas áreas, como o $A_i B_i$ (UNITED NATIONS, 1956; MADEIRA; SIMÕES, 1972), o método “quase-componente” (HAKKERT, 1985), e o ProjPeq (JANNUZZI, 2007), não captam a inter-relação existente entre pequenas áreas. Para tal, o mais apropriado seria uma abordagem multirregional, capaz de “espelhar as especificidades das trocas populacionais interestaduais, inter-regionais e, particularmente, intra-regional” (BAENINGER; CUNHA, 1996). Como existem diferenças socioeconômicas significativas ao longo do espaço que tendem a influenciar a dinâmica demográfica de cada área, a previsão populacional a partir do município de residência é justificável, sobretudo quando o componente migratório possui um papel fundamental na dinâmica de crescimento de áreas menores. A partir de previsões populacionais multirregionais, consideram-se as heterogeneidades demográficas espaciais e, supostamente, melhoram-se a qualidade e a precisão das previsões, já que o método assegura consistência entre as projeções dos vários níveis de agregação. Além de considerar as especificidades demográficas de cada região ao permitir que ela tenha suas próprias taxas de crescimento natural e migração, o método multirregional também garante que a projeção das regiões menores será compatível e consistente com a projeção esperada

para a região maior. Nesse sentido, a modelagem multirregional pode ser entendida como a extensão dinâmica e multidimensional do método das componentes.²

Outro desafio na área de projeções refere-se à trajetória futura das componentes que afetam o crescimento demográfico. Para decidir o caminho a ser seguido pela fecundidade, mortalidade e migração, demógrafos têm adotado quatro alternativas (BOOTH, 2006). A primeira consiste em extrapolar tendências históricas e assumir que o comportamento futuro será uma continuação da tendência passada usando metodologias de séries temporais (LEE; CARTER, 1992; LEE; TULJAPURKAR, 1994). A segunda alternativa consiste em definir hipóteses e trajetórias futuras para o período da projeção utilizando o conhecimento e intuição de especialistas. Tipicamente definem-se cenários de crescimento baixo, médio e alto e, a partir daí, assume-se que o valor esperado estará dentro do intervalo subjetivamente definido (UNITED NATIONS, 1973; WALDVOGEL et al., 2003; IBGE et al., 2004). A terceira consiste em modelar o comportamento dos componentes por meio do uso de covariáveis exógenas relacionadas. Esse método, entretanto, só se torna superior aos demais quando a capacidade de prever o comportamento futuro das covariáveis relacionadas é maior do que a capacidade de prever o comportamento das variáveis demográficas propriamente ditas (KEYFITZ, 1981; 1982). Por fim, a quarta alternativa consiste em considerar a incerteza por meio da simulação de trajetórias futuras (PFLAUMER, 1988; LEE, 1992, 1998; LEE; TULJAPURKAR, 1994; FÍGOLI, 1999). O método consiste em simular trajetórias da população a partir da seleção aleatória de parâmetros e variáveis relacionadas ao crescimento populacional dentro de uma dada distribuição estatística. Para tanto, as estatísticas vitais são geradas pelo método de simulação de Monte Carlo, conforme uma distribuição de probabilidade assumida. De acordo com Fígoli (1999, p. 28), suas vantagens são: “a de ser conceitualmente simples, a de fornecer os resultados a nível da distribuição por idade da população e a de permitir o estabelecimento do intervalo de confiança para todos os resultados da projeção.” Esta abordagem aproxima-se da metodologia bayesiana, tida como a vanguarda dos métodos de previsão populacional (GUIMARÃES, 2014; WISNIOWSKI et al., 2015).

Este artigo introduz um método de previsão populacional capaz de integrar as especificidades locais da dinâmica demográfica municipal, combinando equações lineares bayesianas, variabilidade espacial e métodos de simulação Monte Carlo. O principal resultado são intervalos de confiança para a população do estado e dos municípios paulistas que levam em conta a incerteza futura dos parâmetros associados à dinâmica demográfica. O

² A abordagem multiestado foi primeiramente introduzida por Andrei Rogers (1968) para estudar o crescimento inter-regional e, posteriormente, desenvolvida por Rogers (1971, 1975, 1995) e Schoen (1988, 2006). Na projeção multiestado, a transição entre diferentes “estados” (ex. regiões, estado conjugal, participação na força de trabalho, condição de saúde) pode ser modelada de acordo com a especificidade de cada estado (BOOTH, 2006). Nos Estados Unidos e na Europa há várias aplicações deste método, mas no caso brasileiro a abordagem multiestado é menos comum. Poucas exceções são os estudos de Machado (1993), em nível nacional, e Fígoli et al. (2003), em nível estadual, para projetar a população por idade e sexo utilizando dados do Censo Demográfico de 1991. Muniz (2012) adotou a abordagem multiestado para projetar e simular o crescimento das populações das classes baixa, média e alta e para mensurar como o crescimento destes três grupos poderia afetar os níveis de desigualdade de renda no Brasil.

método desenvolvido utiliza os municípios paulistas, mas poderia também ser aplicado a outros níveis de agregação geográfica e cortes populacionais por idade, sexo, raça, escolaridade, etc.

Objetivos

Este artigo tem quatro objetivos interligados e interdependentes. O primeiro é construir matrizes multirregionais de crescimento populacional para os municípios paulistas, levando em conta as trocas migratórias intermunicipais e o crescimento natural específico de cada município. O segundo objetivo é descrever padrões e tendências de trocas migratórias para os municípios paulistas utilizando ferramentas de visualização e análise espaciais de dados. Essa parte do artigo identifica grupos de municípios com maior e menor crescimento demográfico e distingue o impacto da migração e do crescimento vegetativo nos municípios paulistas a partir da matriz de crescimento multirregional construída. O terceiro objetivo caracteriza-se pela identificação, construção e estimação de modelos econométricos preditivos das probabilidades de migração intermunicipal e de crescimento natural (fecundidade e mortalidade). Estes modelos mensuram a associação entre a dinâmica migratória e as taxas de crescimento natural passadas e futuras utilizando métodos bayesianos. Por fim, o quarto objetivo é demonstrar uma metodologia simples e integrada para projetar (e prever) a população dos municípios de São Paulo, considerando a incerteza inerente ao processo de projeção. Os objetivos descritos são desenvolvidos em quatro módulos marcados por metodologias integradas e contribuições específicas.

Dados e métodos

Módulo 1. Construção da matriz de projeção multirregional

A matriz de projeção multirregional integra a fecundidade, mortalidade e mobilidade de diferentes municípios em um instrumento analítico único. Seguindo as ideias de Rogers (1968), a matriz do sistema de projeção multirregional possui dois componentes: os elementos na diagonal principal da matriz representam o efeito combinado da fecundidade, mortalidade e emigração de cada município; e os elementos fora da diagonal principal representam as probabilidades de imigração intermunicipal dos municípios j para os municípios i . Simbolicamente, para um intervalo de projeção de cinco anos, a matriz de crescimento inter-regional é representada pela relação:

$$w_{\text{proj}}^{t+5} = (B - D + P)^* w^t = G * w^t \quad (1)$$

Onde:

w^t = vetor-coluna cujos elementos denotam a população da região i no tempo t ;

B, D = matrizes-diagonal cujos elementos são não negativos e representam as taxas brutas de natalidade e mortalidade dos municípios;

P' = transposta da matriz de probabilidade de migrar da região de origem para a região de destino. A matriz quadrada P é composta por elementos p_{ij} , que descrevem a proporção de pessoas que durante um momento específico migram da região i para a região j ;

$G = (B - D + P')$ = matriz de crescimento inter-regional.

Os elementos de P' são calculados utilizando-se a informação sobre o migrante de “data fixa” contida na amostra dos Censos Demográficos de 1991 e 2000. A pergunta sobre o município de residência anterior, cinco anos antes do censo, permite o cálculo das probabilidades de migração intermunicipal (CUNHA, 2005). Para isso, basta estimar uma matriz de fluxos origem-destino e dividir cada um dos elementos da mesma pela soma dos elementos da linha correspondente.³ Já os elementos das matrizes B e D serão combinados em uma única matriz-diagonal para representar o crescimento natural e a emigração dos municípios selecionados na análise. Cada elemento da diagonal principal da matriz G representa o efeito combinado da fecundidade, mortalidade e emigração da região i . O cálculo destes elementos é feito de maneira residual, considerando a diferença entre a população projetada pela matriz P' , $W_{proj}^{t+5,i}$, e a população observada em $W^{t+5,i}$. O crescimento referente à migração dos menores de cinco anos e dos filhos de migrantes também está incluído nos valores da diagonal principal da matriz de crescimento multirregional.

Para ilustrar o funcionamento do sistema de crescimento multirregional, considera-se a matriz fictícia de probabilidades de migração entre São Paulo e o resto do Brasil, representada na Tabela 1.

TABELA 1
Matriz de transição de fluxos migratórios
Estado de São Paulo e resto do Brasil – 1985-1990

Origem	Destino		População total em 1985	População total em 1990
	São Paulo	Resto do Brasil		
São Paulo (P_{ii}, P_{ij})	26.528 (0,9373)	1.775 (0,0627)	28.303 (1,0000)	30.783
Resto do Brasil (P_{ji}, P_{jj})	1.345 (0,0127)	104.580 (0,9873)	105.925 (1,0000)	113.308

Em milhões de pessoas

Fonte: Probabilidades para São Paulo adaptadas a partir de Rogers (1968).

Utilizando-se os dados da Tabela 1 para compor a matriz P' , sem considerar o efeito combinado da fecundidade, mortalidade e emigração, a população em 1990 seria dada por:

$$W_{proj}^{1990} = P' * W^{1985} = \begin{bmatrix} 0,9373 & 0,0127 \\ 0,0627 & 0,9873 \end{bmatrix} \begin{bmatrix} 28.303 \\ 105.925 \end{bmatrix} = \begin{bmatrix} 27.874 \\ 106.354 \end{bmatrix} \quad (2)$$

Os valores da matriz P' mostram que a probabilidade de estar vivendo em São Paulo em 1985 e em 1990 era de 93%, e a probabilidade de ter imigrado de alguma parte do

³ Como os dados de migração resultam de estimativas amostrais, municípios menores tenderão a apresentar probabilidades temporalmente instáveis, sobretudo ao desagregá-las por subgrupos ainda menores, por sexo e idade. Tal variabilidade seria resolvida utilizando-se áreas maiores (ex. microrregiões) ou suavizando-se as probabilidades a partir das taxas de migração das áreas do entorno. Estes procedimentos, no entanto, não serão adotados neste artigo, já que o componente migratório é nulo em 93% dos municípios ou representa uma parcela muito pequena do crescimento demográfico observado.

Brasil entre 1985 e 1990 para São Paulo era de 1,27%. Nota-se que, como a matriz P' não leva em conta o crescimento vegetativo do período e nem a migração dos menores de cinco anos, a população projetada é menor do que a de fato observada. Após ajustar os valores da diagonal principal para que a população observada em 1990 seja atingida, tem-se:

$$w^{1990} = G * w^{1985} = \begin{bmatrix} 1,0401 & 0,0127 \\ 0,0627 & 1,0529 \end{bmatrix} \begin{bmatrix} 28.303 \\ 105.925 \end{bmatrix} = \begin{bmatrix} 30.783 \\ 113.308 \end{bmatrix} \quad (3)$$

A matriz G pode então ser utilizada para realizar previsões populacionais futuras, medir a contribuição dos imigrantes para o crescimento específico de cada região, analisar populações estáveis e calcular a sensibilidade do crescimento populacional a cada um dos elementos da matriz. Este exemplo simplificado ilustra, portanto, como a mortalidade, fecundidade e mobilidade inter-regionais podem ser combinadas para descrever o crescimento inter-regional, tornando explícito o papel das trocas migratórias.

A matriz de crescimento para projetar o tamanho futuro dos municípios paulistas tem estrutura similar à descrita acima, mas o número de linhas e colunas da matriz G é igual ao resto do Brasil mais os 572 municípios existentes no estado em 1991. Os 73 novos municípios criados durante a década de 1990 foram reagrupados às suas fronteiras políticas iniciais para permitir a comparabilidade geográfica entre os dados dos Censos de 1991 e 2000. A matriz de crescimento gerada em 1991 (G^{1991}) utiliza os dados dos Censos Demográficos de 1991 e 1980, enquanto a matriz de crescimento inter-regional de 2000 (G^{2000}) emprega os dados de 1991 e 2000.

Módulo 2. Descrição espacial de padrões e tendências de crescimento

Por se tratar de muitos municípios, os elementos da matriz de crescimento G serão visualizados e analisados por meio de técnicas de geoprocessamento. Após a compatibilização temporal das fronteiras políticas de cada município, mapearam-se as seguintes variáveis:

- (a) porcentagem do crescimento municipal devido à entrada de migrantes, em 1991 e 2000;
- (b) porcentagem do crescimento devido à fecundidade, mortalidade e emigração em 1991 e 2000 – elementos da diagonal principal;
- (c) diferença temporal entre as porcentagens descritas em (a) e (b).

Neste módulo conduzimos dois tipos de análise: a primeira indica se as áreas de maior e de menor crescimento tendem a estar próximas entre si, formando agrupamentos de crescimento demográfico (MIRON, 1984; ANSELIN, 1995; FOTHERINGHAM; BRUNSDON, 1999; ANSELIN, 2003). Muniz (2009), por exemplo, utilizou esta técnica para identificar padrões de distribuição espacial das taxas de fecundidade nos municípios brasileiros entre 1980 e 2000, enquanto Jakob (2011) e Cunha e Jakob (2010) a utilizaram para identificar padrões de aglomeração espacial de variáveis socioeconômicas por setores censitários na Baixada Santista e na Região Metropolitana de Campinas.

O segundo tipo de análise consiste no cálculo da estatística de correlação espacial conhecida como I de Moran (MORAN, 1950; CLIFF; ORD, 1981). Esta estatística demonstra o

nível geral de correlação espacial das variáveis municipais descritas em (a), (b) e (c) dentro da distância mínima necessária para garantir a existência de pelo menos um município vizinho (aproximadamente 100 km). A preparação, descrição e análise espacial dos dados foram feitas com o auxílio dos programas computacionais Geoda (ANSELIN et al., 2006) e ArcGis (ESRI, 2012).

Módulo 3. Modelagem da migração intermunicipal e do crescimento natural

O comportamento migratório de cada município paulista é modelado a partir da correlação entre migração passada e futura. Espera-se que a probabilidade passada de migração seja a melhor preditora da probabilidade de se migrar no futuro (GIROSI; KING, 2008). Estimam-se regressões lineares bayesianas para garantir que os coeficientes resultem da própria distribuição observada e não de supostas variações amostrais (como na inferência estatística frequentista). Os coeficientes obtidos poderiam então ser utilizados para prever as probabilidades mínimas e máximas de migração intermunicipal dentro de um intervalo “crível” de confiança estatística de 95%.

Os elementos da diagonal principal da matriz de crescimento inter-regional também serão modelados de acordo com a tendência do crescimento natural entre 1991 e 2000. Estes modelos tendenciais, ainda que simplistas, são factíveis para descrever o crescimento vegetativo municipal, sendo também os mais plausíveis na avaliação de 23 dos 30 especialistas brasileiros consultados em uma pesquisa de opinião realizada por Jannuzzi (2007). Os coeficientes bayesianos subsidiam o módulo de projeção e previsão populacional descrito a seguir.

Módulo 4. Previsão multirregional da população dos municípios paulistas

As previsões populacionais possuem dois tipos de incerteza. A primeira está associada à amostra censitária utilizada para calcular as células da matriz de crescimento. Em áreas pequenas pode haver grande instabilidade das probabilidades calculadas. Já a segunda incerteza está atrelada ao valor futuro das células da matriz de crescimento. Para lidar com estes dois tipos de incerteza utilizamos métodos bayesianos para incorporar a própria distribuição de probabilidade das células da matriz de crescimento, permitindo que as mesmas variem segundo o desvio padrão dos elementos específicos dentro e fora da diagonal principal da matriz de projeção.

Em um primeiro momento utiliza-se a matriz gerada em 1991 (G^{1991}) para se projetar a população até 2000. Esse cenário assume que as condições demográficas permanecem as mesmas ao longo da década, mantendo-se os elementos da matriz G^{1991} constantes. Esse exercício, apesar de irrealista, é instrutivo para se avaliar como a população de cada município e do estado crescerá se as condições demográficas do passado permanecessem as mesmas.

Em um segundo momento, considera-se a incerteza futura na geração das projeções populacionais entre 1991 e 2000, utilizando amostras aleatórias de parâmetros de uma dada distribuição para explorar o comportamento de processos complexos. O método cria milhares de cenários de projeção demográfica por meio de mudanças semialeatórias nos elementos da matriz de crescimento, permitindo a obtenção de resultados intervalares de projeção, ou previsões populacionais. Os coeficientes e variáveis preditas para cada período de projeção são gerados conforme a distribuição prévia de probabilidade das variáveis consideradas pelos modelos de regressão propostos.

Considere-se que, para projetar a população de 1990 para 1995 e de 1995 para 2000, duas matrizes de crescimento multirregional são necessárias. A primeira matriz, G^{1991} , ao ser multiplicada por um vetor-coluna contendo a população dos municípios paulistas em 1990, produz a população projetada em 1995. Realizada desta forma, a projeção assume que o crescimento vegetativo relativo e as probabilidades de mobilidade entre 1990 e 1995 são iguais aos observados entre 1986 e 1991. Para flexibilizar este pressuposto e adicionar alguma incerteza à projeção futura, um elemento de aleatoriedade é adicionado aos valores dos elementos fora (p_{ij}^{1991}) ou dentro (c_i^{1991}) da diagonal da matriz de crescimento calculando-se:

$$p_{ij}^{1991} u = p_{ij}^{1991} \varpi, \text{ sendo } \varpi \sim N(1, \sigma) \text{ e aleatório} \quad (4)$$

$$c_i^{1991} u = c_i^{1991} \varphi, \text{ sendo } \varphi \sim N(1, \delta) \text{ e aleatório} \quad (5)$$

Onde: σ e δ são os desvios padrão dos p_{ij}^{1991} ou c_i^{1991} elementos observados na matriz G^{1991} .⁴ Realizando-se a operação em (4) e (5) por cerca de 1.000 vezes, são geradas distribuições de $p_{ij}^{1991} u$ e $c_i^{1991} u$ aproximadamente normais com médias p_{ij}^{1991} e c_i^{1991} para cada elemento fora ou dentro da diagonal principal da matriz G^{1991} . Para gerar matrizes de projeção para o período 1990-1995, combinam-se então as matrizes aleatórias, independentes e simultâneas das distribuições de $p_{ij}^{1991} u$ e $c_i^{1991} u$. Como resultado deste exercício, 1.000 matrizes de crescimento e, conseqüentemente, 1.000 projeções demográficas probabilísticas são geradas para o período seguinte. Estas matrizes são então novamente multiplicadas pelas várias populações projetadas em 1995 para se obter a previsão populacional para 2000.⁵ Medem-se os erros das projeções e da previsão populacional comparando-as com as populações paulistas municipais observadas no Censo de 2000, que servem como referencial para avaliar a qualidade da previsão média realizada de 1991 para 2000.

⁴ Esse pressuposto garante que a variabilidade de cada elemento da matriz de crescimento G^{1991} será proporcional ao desvio padrão das probabilidades de migração municipal observadas no estado de São Paulo. A distribuição de frequência dos elementos ϖ e φ encontra-se no Anexo 3.

⁵ Os elementos das matrizes de crescimento incluindo incerteza, $G^{1991} u$ e $G^{1996} u$, são gerados a partir da matriz do ano base, G^{1991} . Em 1980, a informação sobre migrante de data fixa não estava disponível e, por isso, a construção de uma matriz de crescimento G^{1980} não é possível.

Resultados

Descrição das matrizes inter-regionais de crescimento demográfico

As matrizes de projeção multirregional consideram as trocas migratórias intermunicipais e o crescimento natural específico de cada município. As primeiras 572 linhas e colunas correspondem aos municípios existentes em 1991 e 2000 (compatibilizados para 1991) e a 573ª linha/coluna representa as trocas migratórias entre os municípios paulistas e o resto do Brasil. Para a matriz origem-destino de 1991, o maior valor da diagonal (crescimento natural + emigração) foi igual a 1,397. Isso indica que houve um aumento populacional de 39,75% no município de Santana de Parnaíba entre 1986 e 1991. O menor valor da diagonal foi observado no município de Santa Mercedes (0,658). Esse valor indica que 65,8% da população observada em 1986 neste município se encontrava neste mesmo município em 1991. Já no que se refere aos valores fora da diagonal, do total de células presentes na matriz (328.329), cerca de 7% (23.181) possuíam valores diferentes de zero. Outras medidas descritivas das células contidas dentro e fora da diagonal principal das matrizes de origem-destino de 1991 e 2000 são apresentadas na Tabela 2.

TABELA 2
Descrição dos elementos das matrizes inter-regionais de origem-destino dos municípios
Estado de São Paulo –1991-2000

Localização das células da matriz de crescimento	Matriz origem-destino	
	1991	2000
Diagonal principal (crescimento natural + emigração)		
Número de células	573	573
Média	0,9345	0,9243
Desvio padrão	0,0909	0,0996
Mediana	0,9462	0,9324
Máximo	1,3976	1,1852
Mínimo	0,6588	0,4220
Fora da diagonal principal (probabilidades de imigração)		
% de células diferentes de zero	0,0706	0,0825
Média	0,0027	0,0024
Desvio padrão	0,0063	0,0057
Mediana	0,0006	0,0005
Máximo	0,1665	0,1300
Mínimo	0,0000	0,0000

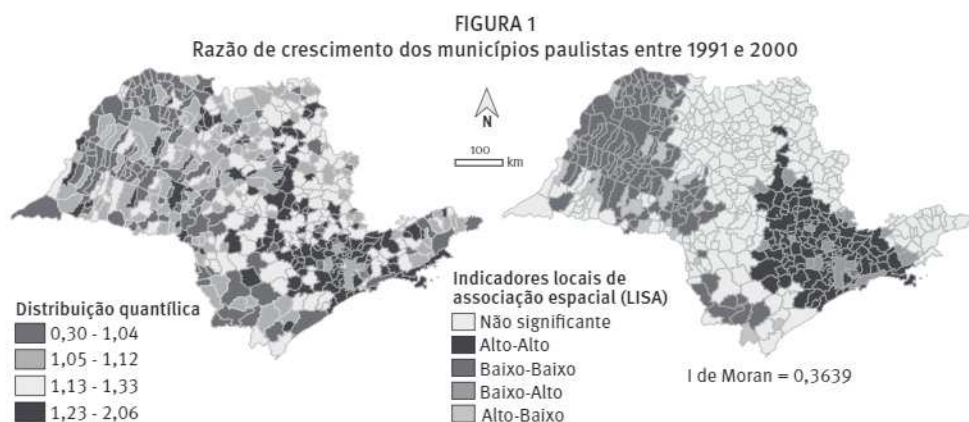
Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Nota: As medidas descritivas para os elementos fora da diagonal principal excluem valores iguais a zero.

Descrição espacial de padrões e tendências de crescimento

O referenciamento espacial permite identificar os grupos de municípios com maior e menor crescimento demográfico e distingue o impacto da migração e do crescimento vegetativo nos municípios paulistas.

A Figura 1 localiza os municípios paulistas que tiveram maior e menor crescimento demográfico entre 1991 e 2000, evidenciando que mais de 50% deles cresceram entre 5% e 22% em dez anos. O mapa à esquerda mostra que, enquanto a capital paulista teve crescimento demográfico entre 5% e 12%, os municípios no seu entorno cresceram entre 23% e 106% no período analisado. Fenômeno similar aconteceu no entorno do município de Campinas.



Fonte: IBGE. Censos Demográficos de 1991 e 2000.

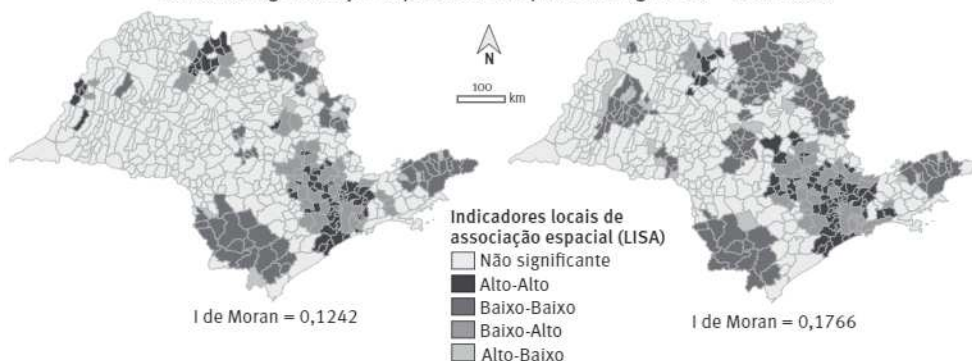
Nota: Padrão de aglomeração baseia-se na distância mínima entre centroides municipais (aprox.100 km).

O mapa à direita na Figura 1 mostra que as mesorregiões de maior crescimento demográfico foram a Macrometropolitana Paulista, a Metropolitana de São Paulo, Piracicaba, norte de Itapetininga, Campinas e sudeste da mesorregião de Bauru. Municípios classificados na categoria “alto-alto” evidenciam crescimento populacional elevado, sendo que os valores médios do entorno (em uma distância média aproximada de 100 km) também são altos. Por outro lado, regiões incluídas na categoria “baixo-baixo” (Araçatuba, São José do Rio Preto, Itapetininga) são municípios nos quais a razão de crescimento apresentou valores baixos. Os valores “baixo-alto”, como no município de São Paulo por exemplo, correspondem a locais nos quais a razão de crescimento é baixa, enquanto o crescimento do entorno é alto. O mapa, entretanto, não separa o peso do componente migratório do crescimento natural na composição do crescimento total dos municípios paulistas. Para isso apresenta-se a Figura 2.

A Figura 2 localiza regiões onde houve maior participação do componente migratório na composição das populações dos municípios paulistas. A Região Metropolitana de São Paulo e alguns municípios na região leste da mesorregião de São José do Rio Preto são as localidades onde o número de migrantes é o mais elevado do estado, enquanto Ribeirão Preto, Itapetininga e Vale do Paraíba Paulista são áreas com baixo percentual de migrantes em sua composição. Os municípios que tinham o maior número de migrantes na composição da sua população total eram Luís Antônio (43%), Itupeva (39,4%) e Bady Bassit (37,12%). A capital do estado possuía 6% de migrantes na sua população. A mesorregião do Litoral Sul Paulista também merece destaque pelo elevado número de migrantes em sua composição populacional, sobretudo nos municípios de Praia Grande, Mongaguá, Itanhaém e Peruíbe.

A Figura 3 mostra os municípios nos quais a participação de migrantes na composição populacional mais cresceu entre 1991 e 2000. Municípios classificados nas categorias “baixo-baixo” e “alto-alto” representam áreas nas quais o percentual de migrantes na população em 2000 foi maior ou igual àquele observado em 1991. Os municípios no entorno da capital e de Campinas tiveram entre 30% e 300% de aumento na quantidade de migrantes compondo suas populações. Os municípios nos quais a participação de migrantes mais aumentou foram Queiroz (379%), Alfredo Marcondes (191%), Conchas (187%), Águas de São Pedro (164%), Sete Barras (160%) e Mongaguá (147%).

FIGURA 2
Padrão de aglomeração espacial do componente migratório – 1991-2000



Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Nota: Padrão de aglomeração espacial baseia-se na distância mínima entre centroides municipais (aprox. 100km).

FIGURA 3
Variação da participação de migrantes entre 2000 e 1991



Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Nota: Padrão de aglomeração espacial baseia-se na distância mínima entre centroides municipais (aprox. 100 km).

O mapa à direita da Figura 3 mostra o padrão de aglomeração espacial dos municípios nos quais o percentual de migrantes mais cresceu entre 1991 e 2000, evidenciando

um aumento na participação de migrantes na população na região localizada entre os municípios de São Paulo e Campinas e naqueles pertencentes às microrregiões de Avaré e Botucatu. Já na microrregião de São João da Boa Vista e arredores houve redução da participação de migrantes na composição populacional de seus municípios.

Modelagem da migração intermunicipal e do crescimento natural

Para prever as probabilidades de migração intraestadual – células fora da diagonal principal da matriz de crescimento inter-regional – especificou-se o seguinte modelo linear generalizado:

$$p_{ij}^{2000} = \beta_0 + \beta_1 * p_{ij}^{1991} + \varepsilon_{ij} \quad \text{sendo } p_{ij}^{2000} \sim \Gamma(\kappa; \theta) \tag{6}$$

Onde: p_{ij} é a proporção de pessoas migrando da região i para a região j . O modelo tem distribuição gama altamente assimétrica a direita, com p_{ij}^{2000} e parâmetros de forma κ e escala θ maiores do que zero; ε_{ij} refere-se ao componente residual do modelo. Os coeficientes estimados pelo método bayesiano, com função gama de máxima verossimilhança, estão na Tabela 3.

TABELA 3
Coeficientes estimados pelo modelo bayesiano linear generalizado, com função gama de máxima verossimilhança, para a probabilidades de migração intermunicipal Estado de São Paulo – 1991-2000

Variável	Descrição	Média	Desvio-padrão	Erro-padrão da média posterior (MCSE)	Média posterior da regressão bayesiana (intervalo crível de 95%)
p_{ii}^{1991}	Probabilidades de migração intermunicipais em 1991	0,8379	0,0174	0,000211	(0,8045; 0,8722)
β_0	Intercepto da regressão	0,00099	9,14E-06	1,30E-07	(0,00097; 0,00101)

Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Nota: As estimativas utilizam 27.089 observações diferentes de zero fora da diagonal principal da matriz de crescimento inter-regional. O modelo de regressão bayesiano utilizou distribuições prévias (*priors*) não informativas. A taxa de aceitação da cadeia de Markov Monte Carlo foi de 37,59% e a taxa média de eficiência foi de 15,42%. Estas taxas indicam boa mistura de cadeia de Markov. Empregou-se amostragem Metropolis-Hasting com passeio aleatório para a média. Outras medidas de ajuste do modelo bayesiano encontram-se no Anexo 2.

Os resultados mostram que a probabilidade média esperada de migração, independentemente daquela observada em 1991, é praticamente zero (0,099%). Levando-se em conta a probabilidade de migração passada, a probabilidade esperada para 2000 altera-se consideravelmente. Quando, por exemplo, em 1991 a probabilidade de migração é igual a 0,01 (ou 1%), em 2000 ela passa a ser de 0,94% [ou 0,00099+0,8379*0,01]. Caso utilizássemos os coeficientes médios estimados para prever as probabilidades de migração em 2010, teríamos um modelo do tipo:

$$\widehat{p_{ij}^{2010}} = 0,00099 + 0,8379 * \widehat{p_{ij}^{2000}} \tag{7}$$

Sendo $\widehat{p_{ij}^{2010}}$ o vetor das probabilidades previstas de imigração intermunicipal entre 1995 e 2000 geradas a partir da equação (6) e $\widehat{p_{ij}^{2000}}$ as probabilidades previstas de imigração intermunicipal entre 2005 e 2010. A incerteza futura sobre o valor destas probabilidades

preditas é refletida pela variação dos coeficientes bayesianos estimados, contidos no intervalo crível de 95% de confiança estatística.

Já para a diagonal principal, que representa o crescimento natural e a emigração ocorrida entre 1995 e 2000, o modelo preditivo possui a seguinte especificação:

$$c_i^{2000} = a_0 + a_i * c_i^{1991} + \varepsilon_i \quad \text{sendo} \quad c_i^{2000} \sim N(\mu; \sigma^2) \tag{8}$$

Onde c_i são os valores da diagonal principal da matriz de crescimento G para o município i em 1991 ou 2000; e ε_i é termo de erro da regressão. Os valores de c_i^{2000} e c_i^{1991} são calculados seguindo a metodologia descrita no módulo 1. Já os coeficientes a_0 e a_i (Tabela 4) são estimados utilizando-se métodos bayesianos com função de máxima verossimilhança linear normal.⁶

TABELA 4
 Coeficientes estimados pela regressão bayesiana, com função de máxima verossimilhança linear normal, para o crescimento natural e emigração intermunicipal
 Estado de São Paulo – 1991-2000

Variável	Descrição	Média	Desvio padrão	Erro padrão da média posterior (MCSE)	Média posterior da regressão bayesiana (Intervalo crível de 95%)
c_i^{1991}	Crescimento demográfico natural e emigração entre 1986 e 1991	0,8200	0,0261	0,000847	(0,7692; 0,8707)
a_0	Intercepto da regressão	0,1630	0,0245	0,000796	(0,1156; 0,2106)

Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Nota: As estimativas baseiam-se em 560 observações da diagonal principal da matriz de crescimento inter-regional (13 observações foram eliminadas da estimativa por terem distância de Cook) 1 ou serem pontos influentes). O modelo de regressão bayesiano utilizou distribuições prévias (*priors*) não informativas. A taxa de aceitação da cadeia de Markov Monte Carlo foi 100% e a taxa média de eficiência foi de 34,6%. Estas taxas indicam boa mistura de cadeia de Markov. Empregou-se amostragem Gibbs tanto para a variância quanto para a média. Outras medidas de ajuste do modelo bayesiano encontram-se nos Anexos 2 e 2.1.

Os coeficientes médios *a posteriori* são próximos daqueles obtidos a partir da abordagem frequentista – o modelo final “explica” quase 70% da variabilidade do crescimento natural municipal no estado de São Paulo em 2000 ($R^2_{ajustado} = 0,66$). A principal vantagem da abordagem bayesiana, no entanto, é fornecer “intervalos críveis” e com interpretação probabilística para os coeficientes. A probabilidade de o coeficiente estimado para c_i^{1991} estar contido dentro do intervalo [0,77; 0,87] é de 95%. Utilizando a estimativa da média posterior, o modelo mostra que os municípios que cresceram 10% (ou 1,1) entre 1986 e 1991 por causas naturais (e emigração) cresceriam 6,5% (ou 0,163 + 0,82*1,1) entre 1995 e 2000, de acordo com a predição do modelo. Já os municípios cujo crescimento natural entre 1986 e 1991 foi igual a 1% (ou 1,01), não teriam aumento demográfico significativo entre 1995 e 2000. Em suma, o modelo mostra que o crescimento natural (mais emigração) no quinquênio seguinte seria, em média, 0,82% daquele observado nos cinco anos anteriores.

⁶ Restringiu-se a amostra aos casos cujo peso calculado pela regressão robusta (BERK, 1990; ROUSSEEUW; LEROY, 1987; VERARDI; CROUX, 2009) fosse maior do que zero para diminuir a influência de valores extremos na estimativa dos coeficientes.

Previsão multirregional da população dos municípios paulistas

O quarto objetivo desta pesquisa é ilustrar uma metodologia integrada para projetar (e prever) a população dos municípios de São Paulo, levando em conta a incerteza inerente ao processo de projeção. Este módulo da pesquisa introduz o novo método proposto e valida as projeções a partir da comparação com as populações observadas e com os resultados de outro método já consolidado e utilizado pelo IBGE, o $A_i B_i$.

Duas projeções e uma previsão populacional foram feitas para todos os municípios do estado de São Paulo. A primeira projeção utiliza a matriz de crescimento gerada em 1991 para projetar a população de 2000, assumindo estabilidade nos parâmetros da matriz. Essa projeção determinística resulta da multiplicação das populações observadas em 1990 pelo quadrado da matriz de crescimento previamente calculada:

$$W_{proj}^{2000} = G^{1991} * G^{1991} * W^{1990} \quad (9)$$

Onde W_{proj}^{2000} representa um vetor-coluna das i populações municipais projetadas para o ano 2000; G^{1991} corresponde à matriz de crescimento observado de 1986 para 1991; e W^{1990} é o vetor-coluna contendo as i populações observadas de 1990.

A segunda projeção é executada utilizando o método $A_i B_i$, desenvolvido pelas Nações Unidas (UNITED NATIONS, 1956) e pioneiramente utilizado no Brasil por Madeira e Simões (1972). A aplicação do método requer a projeção da área maior (no caso o estado de São Paulo) e as distribuições dos totais populacionais dos municípios. Para projetar a população até 2000 foram tomadas as populações municipais observadas nos censos de 1980 e 1991 e a estimativa populacional para o estado publicada em Waldvogel e Capassi (1999).⁷ Pelo método $A_i B_i$ projetam-se as populações municipais para 2000 da seguinte forma:

$$W_{proj}^{2000,i} = W^{1991,i} + \frac{W^{1991,i} - W^{1980,i}}{W^{1991} - W^{1980}} * (W_{proj}^{2000} - W^{1991}) \quad (10)$$

Como antes, $W^{t,i}$ representa as populações do domínio menor, ou seja, dos municípios i ; $W_{proj}^{t,i}$ corresponde à população projetada do município i no tempo t ; e W representa a população do domínio maior, ou seja, do estado de São Paulo.

Para a previsão probabilística populacional consideraram-se as equações (4) e (5) para a obtenção de 1.000 matrizes de crescimento (e consequentemente 1.000 projeções) geradas a partir do método de Monte Carlo e da variabilidade intraestadual (desvio padrão) dos valores dentro e fora da diagonal principal da matriz de crescimento.⁸

⁷ Na época desta publicação projetava-se que a população do estado de São Paulo seria igual a 36.102.875 pessoas. A população de fato observada foi igual a 36.423.321 indivíduos.

⁸ O Anexo 3 apresenta o programa de Stata utilizado para fazer a extração dos parâmetros de variação da matriz de crescimento e as estatísticas descritivas da distribuição, além de exemplificar o histograma de ω e ϕ para 1.000 extrações aleatórias das distribuições normais assumidas. Por se tratar de um processo aleatório, múltiplas extrações de mesmo tamanho apresentarão distribuições ligeiramente distintas das anteriores e, consequentemente, diferentes intervalos de projeção populacional. A incerteza inerente a este processo, entretanto, não causa variações significativas nas matrizes de crescimento utilizadas. Pelo Teorema Central do Limite, um número maior de extrações apenas torna a distribuição mais próxima da distribuição Gaussiana.

Após combinar os valores gerados dentro e fora da diagonal principal para criar 1.000 matrizes de crescimento, utilizou-se o valor médio das projeções para realizar as comparações de precisão do novo método probabilístico. Entre todos os intervalos de projeção populacional gerados para cada município, constatou-se que 96% deles incluíram o verdadeiro valor observado no Censo Demográfico de 2000. Dos 572 municípios paulistas analisados, 545 tiveram populações previstas dentro dos intervalos produzidos e apenas 27 tiveram populações médias fora do intervalo. Os Anexos 4 e 5 ilustram algumas previsões fora e dentro do intervalo de confiança para populações municipais paulistas.

Para mensurar a precisão das projeções e da previsão populacional média, foram utilizados o erro médio percentual da projeção (EMP), o erro médio percentual em módulo (EMPM) e a razão entre as populações observada e projetada (ROP). Estes três indicadores são definidos da seguinte forma:

$$EMP_t = \frac{\sum_{i=1,n} \left(\frac{W_{proj}^{t,i} - W^{t,i}}{W^{t,i}} \right) \times 100}{n} \quad (11)$$

$$EMPM_t = \frac{\sum_{i=1,n} \left| \left(\frac{W_{proj}^{t,i} - W^{t,i}}{W^{t,i}} \right) \times 100 \right|}{n} \quad (12)$$

$$ROP_t = \frac{\sum_{i=1,n} \frac{W^{t,i}}{W_{proj}^{t,i}}}{n} \quad (13)$$

Segundo Devine e Coleman (apud BRITTO; CAVENAGHI; JANNUZZI, 2010, p. 38), o EMP é uma medida de tendência média e pode ser usada como base para testar a presença de tendência média significativa, enquanto o EMPM é uma medida de precisão que indica o quão “perto” as estimativas chegaram do valor observado, em média (DEVINE; COLEMAN, 2003). Já a ROP indica simplesmente o quão maior ou menor a população observada é em relação àquela projetada. Ela representa a média aritmética simples das razões entre as populações municipais observadas e projetadas. O quadro comparativo entre as três projeções realizadas e as populações observadas encontra-se na Tabela 5.

TABELA 5
Erro médio da projeção demográfica, por método utilizado
Municípios de São Paulo – 2000

Método	EMP ₂₀₀₀	EMPM ₂₀₀₀	População estadual projetada para 2000	ROP ₂₀₀₀
Projeção determinística usando matriz de crescimento demográfico	4,57 (22,54)	11,16 (20,11)	38.020.794	0,98 (0,14)
$A_t B_t$	-0,30 (20,86)	10,70 (17,90)	36.075.827	1,03 (0,14)
Valor médio da previsão probabilística de crescimento populacional	5,25 (22,98)	11,57 (20,53)	38.241.176	0,98 (0,14)

Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Nota: Desvios padrão encontram-se entre parênteses.

Segundo Smith et al. (2001), para um horizonte de projeção de dez anos, o EMPM aceitável para pequenos domínios, como municípios, tem ficado entre 8% e 14%. No Brasil, os métodos de projeção populacional para pequenos domínios considerados de boa precisão têm sido aqueles em que mais da metade dos municípios apresentam erro percentual entre -10% e 10% (JARDIM, 1995; WALDVOGEL; CAPASSI, 1999).

Os resultados obtidos são compatíveis com estes padrões. Apesar de a população estadual projetada pelo $A_i B_i$ estar mais próxima daquela de fato observada no Censo de 2000 (36.423.321), o erro médio apresentado pelos outros métodos encontra-se dentro dos limites aceitáveis. O $A_i B_i$ apresenta os menores erros médios (EMP e EMPM), mas todos os métodos possuem precisão aceitável e compatível com aquela sugerida pela literatura. A coluna ROP mostra que, em média, a população dos municípios projetada pelo $A_i B_i$ é 3% maior do que a observada. Já os métodos determinístico e probabilístico tendem a projetar populações municipais que são, em média, 98% do tamanho da população de fato observada.⁹ O Gráfico 1 (à esquerda) apresenta a distribuição de frequência dos erros percentuais da projeção e demonstra que todos os métodos registram erro médio em torno de zero. A imagem à direita do Gráfico 1 mostra a distribuição dos erros absolutos dos municípios mais problemáticos, cuja diferença populacional *ex-post* foi superior a 10 mil pessoas.

A Figura 4 complementa a análise espacial do erro das projeções, identificando municípios cujo crescimento foi discrepante daquele encontrado para os demais.¹⁰ O cartograma mostra a localização aproximada dos municípios cuja razão entre a população observada e a projetada foi sub ou sobrestimada. O padrão de pontos extremos é similar entre os dois métodos, indicando que os municípios nos quais as populações são sub ou sobrestimadas são os mesmos, independentemente do método de projeção utilizado (o coeficiente de determinação entre os erros das duas projeções foi igual a 0,91).

⁹ No caso da projeção probabilística, especula-se que a população total média do estado seja maior do que a observada por causa dos valores extremos projetados, gerados pelo fato de as projeções incluírem todos os valores preditos, e não apenas aqueles com 95% de confiança estatística. Pelo método probabilístico, a população prevista para o estado estaria entre 24.531.361 e 54.007.595. Este intervalo de confiança seria mais estreito adotando-se níveis de confiança menores (ex. contidos entre aproximadamente dois desvios padrão da média populacional prevista).

¹⁰ Os mapas para a projeção determinística apresentaram distribuição similar à da projeção probabilística.

GRÁFICO 1
Erros percentuais e absolutos de projeção populacional
Municípios de São Paulo – 2000

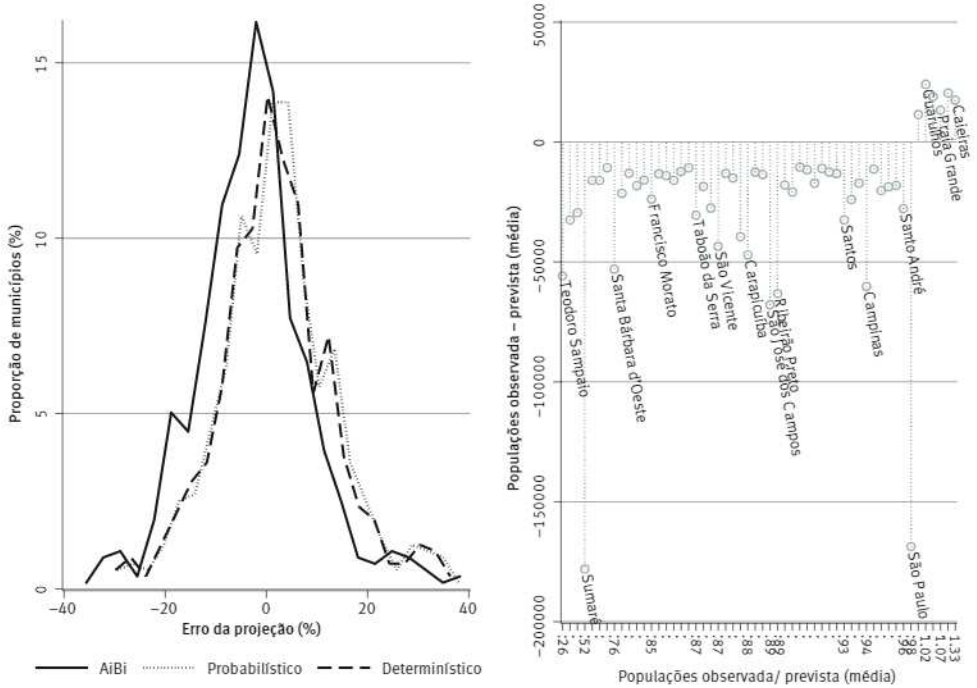
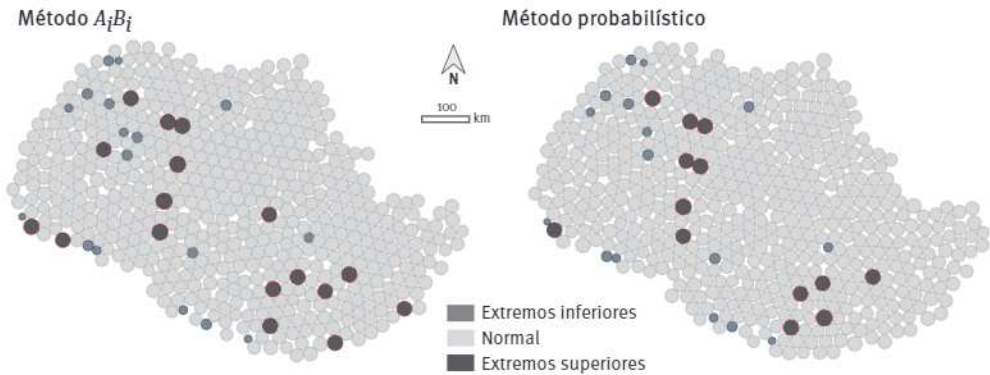


FIGURA 4
Cartograma da razão entre as populações observada e projetada
Municípios de São Paulo – 2000



Discussão e contribuições

No Brasil, a estratégia metodológica dominante para realizar projeções em pequenas áreas consiste em utilizar o método das componentes para projetar a população para áreas maiores e, então, estimar o número de indivíduos em áreas menores por meio de métodos de ajuste proporcional ou interpolativo – tais como $A_i B_i$. A outra opção, conhecida como modelo ProjPeq, “prescinde de investimentos ou esforços técnicos muito vultosos na produção de projeções de totais populacionais para pequenas áreas” e “não produz necessariamente estimativas futuras mais precisas do que os outros métodos” (JANNUZZI, 2007, p. 111, 130). A metodologia probabilística adotada neste artigo, entretanto, difere dos métodos tradicionalmente utilizados. A primeira diferença é fornecer estimativas populacionais que partem de uma área menor para uma área maior de análise, levando em conta as especificidades regionais de maneira integrada. A segunda diferença é utilizar um método de modelagem estrutural que considera a incerteza atribuída ao comportamento futuro da dinâmica demográfica por meio de métodos bayesianos que utilizam cadeias de Markov e simulações Monte Carlo. Em particular, esta técnica permite a construção de intervalos de confiança estatística para a previsão populacional municipal e para o estado de São Paulo. Os intervalos de confiança apresentados (Anexos 4 e 5), provavelmente, seriam menores caso os componentes de variabilidade dos elementos das matrizes de crescimento fossem segmentados segundo padrões similares de crescimento, conforme aqueles apresentados na Figura 1, por exemplo. Isso implicaria considerar incertezas futuras distintas para municípios de diferentes portes e com diferentes dinâmicas de crescimento, em vez de considerar a variabilidade do estado como um todo para a projeção específica dos municípios. Os resultados dos modelos estimados neste artigo, além disso, ainda precisam ser explorados em previsões populacionais de médio e longo prazos (até 2030, por exemplo). Refinamentos futuros do método proposto deverão, portanto, considerar estas sugestões.

A terceira vantagem do método está na sua simplicidade quanto aos insumos necessários para a sua aplicação. A matriz de crescimento multirregional pode ser construída a partir de apenas três informações disponíveis nos Censos Demográficos de 1991 e 2000: o município de residência anterior, no ano $t-n$; e as distribuições populacionais por município no ano $t-n$ e t . Como o método não requer estimação de taxas de fecundidade e mortalidade para pequenas áreas, sua execução e operacionalização são muito mais simples e diretas do que o coorte-componente ou o ProjPeq.

Em quarto lugar, o método sugerido explícita o papel dos fluxos migratórios na dinâmica de crescimento intermunicipal ao separá-los do efeito combinado da mortalidade e da fecundidade. No caso de pequenas áreas, a mensuração e modelagem dos fluxos migratórios são cruciais, já que parcela do crescimento demográfico municipal atrela-se à dinâmica deste componente. Além disso, a matriz de crescimento multirregional permite a confecção de cenários contrafactuais, simulações e análises de sensibilidade com maior facilidade. Entre as perguntas que a matriz de crescimento populacional ajudaria a responder

destacam-se: qual o efeito marginal da migração e do crescimento vegetativo municipal sobre o crescimento demográfico de São Paulo? Qual seria o crescimento inter-regional e estadual caso os níveis de migração intermunicipal fossem diferentes? Qual seria o nível de intervenção requerido para redirecionar o sistema de crescimento inter-regional para uma distribuição alvo? Estas perguntas podem ser mais facilmente respondidas utilizando a matriz de crescimento inter-regional.

Em quinto lugar, análises de estabilidade e convergência podem ser feitas sem grandes complicações a partir da matriz de crescimento. Pode-se perguntar, por exemplo, quais seriam as consequências de longo prazo de se manter um calendário de crescimento constante, qual a taxa de crescimento intrínseco, estrutura e distribuição da população estável e qual é o número de anos de crescimento até se atingir a estabilidade. Em particular, análises de estabilidade são cruciais, já que elas determinam, em grande parte, o quanto os parâmetros da projeção irão variar no futuro. Willekens (1990, p. 17), por exemplo, argumenta que mais importante do que discutir a precisão de projeções de curto e longo prazos, seria discutir quando um sistema demográfico está próximo da estabilidade ou não, uma vez que isso é o que irá determinar e auxiliar a definição de hipóteses futuras para a projeção populacional.

Em sexto lugar, as projeções produzidas poderão ser combinadas com estimativas municipais alternativas, já que previsões populacionais médias podem gerar melhores resultados do que previsões monometodológicas. Sanderson (1998) e Ahlburg (1998), por exemplo, exploraram as vantagens de se combinarem previsões utilizando diferentes métodos e concluíram que, quando combinados, os modelos econômico-demográficos de países em desenvolvimento produzem estimativas que são tão boas quanto aquelas obtidas pelo método das coorte-componentes. Eles acreditam que a combinação de diferentes previsões populacionais produz resultados mais precisos.

Por fim, as previsões populacionais realizadas por meio do método proposto tiveram um erro *ex-post* aceitável e similar às projeções feitas pelo já consolidado $A_t B_t$, indicando que há potencial para que o método seja adotado em outras regiões e para diferentes níveis de agregação, por idade e sexo por exemplo. Em particular, sugere-se que a restrição da variabilidade dos elementos da matriz de crescimento (ϖ e ϕ) a certas regiões do estado – que apresentaram padrões similares de crescimento demográfico ao longo do espaço – possa vir a aprimorar ainda mais as projeções apresentadas.

Referências

AHLBURG, D. A. Using economic information and combining to improve forecast accuracy in demography. Minneapolis: University of Minnesota, 1998 (Working paper, Industrial Relations Center).

ALHO, J.; SPENCER, B. Statistical demography and forecasting. United States: Springer, 2005.

ANSELIN, L. Local Indicators of Spatial Association – Lisa. *Geographical Analysis*, v. 27, n. 2, p. 93-115, 1995.

- ANSELIN, L. An introduction to spatial autocorrelation analysis with GeoDa. Urbana-Champaign, IL: Spatial Analysis Laboratory (SAL), Department of Agricultural and Consumer Economics, University of Illinois, 2003. Disponível em: <<http://www.uni-kassel.de/~rkosfeld/lehre/spatial/spauto.pdf>>. Acesso em: 4 fev. 2015.
- ANSELIN, L.; SYABRI, I.; KHO, Y. GeoDa: An introduction to spatial data analysis. *Geographical Analysis*, v. 38, n. 1, p. 5-22, 2006.
- BAENINGER, R.; CUNHA, J. M. Migração, dinâmica regional e projeções populacionais. *São Paulo em Perspectiva*, v. 10, n. 2, p. 102-109, 1996.
- BERK, R. A. A primer on robust regression. In: FOX, J.; LONG, J. S. (Ed.). *Modern methods of data analysis*. Newbury Park, CA: Sage, 1990. p. 292-324.
- BOOTH, H. Demographic forecasting 1980 to 2005 in review. *International Journal of Forecasting*, v. 22, n. 3, p. 547-581, 2006.
- BRITO, L.; CAVENAGHI, S.; JANNUZZI, P. M. Estimativas e projeções populacionais para pequenos domínios: uma avaliação da precisão para municípios do Rio de Janeiro em 2000 e 2007. *Revista Brasileira de Estudos de População*, v. 27, n. 1, p. 35-57, 2010.
- CAVENAGHI, S.; POTTER, J.; SCHMERTMANN, C.; ASSUNÇÃO, R. Estimating total fertility rates for small areas in Brazil. In: MEETING OF THE POPULATION ASSOCIATION OF AMERICA. *Proceedings...* Boston, MA: PAA, 2004.
- CLIFF, A. D.; ORD, J. K. *Spatial processes: models and applications*. London: Pion, 1981.
- CUNHA, J. M. Migração e urbanização no Brasil: alguns desafios metodológicos para análise. *São Paulo em Perspectiva*, v. 19, n. 4, p. 3-20, 2005.
- CUNHA, J. M.; JAKOB, A. Segregação socioespacial e inserção no mercado de trabalho na Região Metropolitana de Campinas. *Revista Brasileira de Estudos de População*, v. 27, n. 1, p. 115-139, 2010.
- DEVINE, J.; COLEMAN, C. *People might move but housing units don't: an evaluation of the state and county housing unit estimates*. Washington: Population Division, U.S. Census Bureau, 2003 (Working Paper Series, n. 71).
- DUCHESNE, L. *Proyecciones de poblacion por sexo y edad para areas intermedias y menores*. Santiago: Centro Latinoamericano de Demografia, 1987.
- ESRI – Environmental Systems Research Institute. *ArcGIS Release 10.1*. Redlands, CA, 2012.
- FÍGOLI, M. Intervalo de confiança para projeção de população baseado no método de Monte Carlo: projeção dos beneficiários urbanos da previdência social. *Revista Brasileira de Estatística*, v. 60, n. 213, p. 25-51, 1999.
- FÍGOLI, M. G. B.; WONG, L. L. R.; SAWYER, D. R. T. O.; CARVALHO, J. A. M. Proyección multirregional: aplicación en Brasil y sus unidades federativas (200-2020). *Notas de Población*, Santiago do Chile, v. 76, p. 7-35, 2003.
- FOTHERINGHAM, A.; BRUNSDON, C. Local forms of spatial analysis. *Geographical Analysis*, v. 31, p. 340-358, 1999.
- FREIRE, F. H.; ASSUNÇÃO, R. M. Intervalo de confiança para a taxa de fecundidade total de pequenas áreas. In: XIV ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS. *Anais...* Caxambu, MG: Abep, 1998. Disponível em: <<http://www.abep.nepo.unicamp.br/docs/anais/PDF/1998/a240.pdf>>. Acesso em: 4 fev. 2015.
- GIROSI, F.; KING, G. *Demographic forecasting*. Princeton, NJ: Princeton University Press, 2008.
- GUIMARÃES, R. R. Uncertainty in population projections: the state of the art. *Revista Brasileira de Estudos de População*, São Paulo, v. 31, n. 2, p. 277-290, 2014.

HAKKERT, R. Quasi-component models for small area population projections: examples from the state of São Paulo, Brazil. *International Union for the Scientific Study of Population*. Florence, 1985.

IBGE. *Projeções da população: Brasil e unidades da federação*. Rio de Janeiro, 2013 (Série Relatórios Metodológicos, v. 40). Disponível em: <[ftp://ftp.ibge.gov.br/Projecao_da_Populacao/Projecao_da_Populacao_2013/srm40_projecao_da_populacao.pdf](http://ftp.ibge.gov.br/Projecao_da_Populacao/Projecao_da_Populacao_2013/srm40_projecao_da_populacao.pdf)>. Acesso em: 27 dez. 2016.

_____. *Estimativas da população residente para os municípios e para as unidades da federação brasileiros com data de referência em 1º de julho de 2016*. Rio de Janeiro, setembro de 2016. Disponível em: <<http://biblioteca.ibge.gov.br/visualizacao/livros/liv97868.pdf>>. Acesso em: 27 dez. 2016.

JAKOB, A. Zonas de segregação e vulnerabilidade na Baixada Santista. In: VAZQUEZ, D. A. (Org.). *A questão urbana na Baixada Santista: políticas, vulnerabilidades e desafios para o desenvolvimento*. 1. ed. São Paulo: Editora Universitária Leopoldianum, 2011. p. 65-80.

JANNUZZI, P. M. Cenários futuros e projeções populacionais para pequenas áreas: método e aplicação para distritos paulistanos 2000-2010. *Revista Brasileira de Estudos de População*, v. 24, n. 1, p. 109-136, 2007.

JARDIM, M. L. *Utilização de variáveis sintomáticas para estimar a distribuição espacial de populações: aplicação aos municípios do Rio Grande do Sul*. Belo Horizonte: UFMG/Cedeplar, 1995.

KEYFITZ, N. The limits of population forecasting. *Population and Development Review*, v. 7, n. 4, p. 579-593, 1981.

_____. Can knowledge improve forecasts? *Population and Development Review*, v. 8, n. 4, p. 729-751, 1982.

KEYFITZ, N.; CASWELL, H. *Applied mathematical demography*. New York, NY: Springer, 2005.

LEE, R. D. Stochastic demographic forecasting. *International Journal of Forecasting*, v. 8, n. 3, p. 315-327, 1992.

_____. Probabilistic approaches to population forecasting. *Population and Development Review*, v. 24 (Supplement: Frontiers of Population Forecasting), p. 156-190, 1998.

LEE, R. D.; CARTER, L. Modelling and forecasting the times series of US mortality. *Journal of the American Statistical Association*, v. 87, n. 419, 1992.

LEE, R. D.; TULJAPURKAR, S. Stochastic population forecasts for the United States: beyond high, medium, and low. *Journal of the American Statistical Association*, v. 89, n. 428, p. 1175-1189, 1994.

LESLIE, P. H. On the use of matrices in certain population mathematics. *Biometrika*, v. 33, n. 3, p. 183-212, 1945.

_____. Some further notes on the use of matrices in population mathematics. *Biometrika*, v. 35, n. 3/4, p. 213-245, 1948.

LUTZ, W.; VAUPEL, J.; AHLBURG, D. A. *Frontiers of population forecasting*. New York: Population Council, 1999.

MACHADO, C. *Projeções multirregionais de população: o caso brasileiro (1980-2020)*. Tese (Doutorado) – Centro de Desenvolvimento e Planejamento Regional (Cedeplar), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, 1993.

MADEIRA, J. L.; SIMÕES, C. C. Estimativas preliminares da população urbana e rural segundo as unidades da federação, de 1960/1980 por uma nova metodologia. *Revista Brasileira de Estatística*, v. 33, n.129, p. 3-11, 1972.

MIRON, J. Spatial autocorrelation in regression analysis: a Beginner's Guide. In: GAILE, G. L.; WILLMOTT, C. J. (Ed.). *Spatial statistics and models*. Dordrecht, Holland: D. Reidel Publishing Co., 1984. p. 201-222.

MORAN, P. A. P. Notes on continuous stochastic phenomena. *Biometrika*, v. 37, n. 1/2, p. 17-23, 1950.

MUNIZ, J. O. Spatial dependence and heterogeneity in ten years of fertility decline in Brazil. *Population Review*, v. 48, n. 2, p. 32-65, 2009.

_____. Demographic dynamics of poverty and income inequality: the case of Brazil. *Revista Brasileira de Estudos de População*, v. 29, n. 2, p. 323-348, 2012.

OLIVEIRA, J. C.; ALBUQUERQUE, F. R.; LINS, I. B. *Projeção da população do Brasil por sexo e idade para o período 1980-2050 – Revisão 2004; Estimativas anuais e mensais da população do Brasil e das unidades da federação: 1980-2020; Estimativas das populações municipais*. Rio de Janeiro: IBGE, 2004. Disponível em: <<http://www.ibge.gov.br/home/estatistica/populacao/estimativa2004/metodologia.pdf>>. Acesso em: 4 fev. 2015.

PFLAUMER, P. Confidence intervals for population projections based on Monte Carlo methods. *International Journal of Forecasting*, v. 4, n. 1, p. 135-142, 1988.

ROGERS, A. *Matrix analysis of interregional population growth and distribution*. Berkeley: University of California Press, 1968.

_____. *Matrix methods in urban and regional analysis*. San Francisco: Holden-Day, 1971.

_____. *Introduction to multiregional mathematical demography*. New York: John Wiley, 1975.

_____. *Multiregional demography: principles, methods and extensions*. London: John Wiley, 1995.

ROUSSEEUW, P. J.; LEROY, A. M. *Robust regression and outlier detection*. New York: Wiley, 1987.

SANDERSON, W. Knowledge can improve forecasts: a review of selected socioeconomic population projection models. *Population and Development Review*, v. 24 (Supplement: Frontiers of Population Forecasting), p. 88-117, 1998.

SANTOS, R. O.; BARBIERI, A. F. Projeções populacionais em pequenas áreas: uma avaliação comparativa de técnicas de extrapolação matemática. *Revista Brasileira de Estudos de População*, v. 32 n. 1, p. 139-163, 2015.

SCHOEN, R. *Modeling multigroup populations*. New York: Plenum Press, 1988.

_____. *Dynamic population models*. Dordrecht: Springer, 2006.

SEADE. *Sistema de projeções populacionais para os municípios do estado de São Paulo*. São Paulo, [s/d]. Disponível em: <http://produtos.seade.gov.br/produtos/projpop/pdfs/projpop_metodologia.pdf>. Acesso em: 27 dez. 2016.

SMITH, S.; TAYMAN, J.; SWANSON, D. *State and local population projections: methodology and analysis*. New York: Kluwer, 2001.

STOTO, M. A. The accuracy of population projections. *Journal of the American Statistical Association*, v. 78, n. 381, p. 13-20, 1983.

SZWARCWALD, C. L.; CASTILHO, E. A. Proposta de um modelo para desagregar projeções demográficas de grandes áreas em seus componentes geográficos. *Revista de Saúde Pública*, v. 23, n. 4, p. 269-76, 1989.

UNITED NATIONS. *Manual III. Methods for population projections by sex and age*. New York: Department of International Economic and Social Affairs, 1956.

_____. **The determinants and consequences of population trends: new summary of findings on interaction of demographic, economic and social factors.** New York: Department of International Economic and Social Affairs, 1973.

VERARDI, V.; CROUX, C. Robust regression in Stata. *Stata Journal*, v. 9, n. 3, p. 439-453, 2009.

WALDVOGEL, B. C. Projeção populacional em São Paulo: um método analítico como alternativa. *São Paulo em Perspectiva*, v. 10, n. 2, p. 110-115, 1996.

WALDVOGEL, B. C.; CAPASSI, R. Cenário da população paulista dos anos 90 ao futuro. *São Paulo em Perspectiva*, v. 13, n.1-2, p. 186-195, 1999.

WALDVOGEL, B. C.; FERREIRA, C. E.; YAZAKI, L. M.; GODINHO, R. E.; PERILLO, S. R. Projeção da população paulista como instrumento de planejamento. *São Paulo em Perspectiva*, v. 17, n. 3-4, p. 67-79, 2003.

WILLEKENS, F. **Demographic forecasting: State-of-the-art and research needs.** In: HAZEU, C. A.; FRINKING, G. B. A. (Ed.). *Emerging issues in demographic research.* Amsterdam: Elsevier, 1990. p. 9-66.

WISNIOWSKI, A.; SMITH, P. W. F.; BIJAK, J.; RAYMER, J.; FORSTER, J. Bayesian population forecasting: extending the Lee-Carter method. *Demography*, v. 52, n. 3, p. 1035-1059, 2015.

WHELPTON, P. K. Population of the United States, 1925 to 1975. *The American Journal of Sociology*, v. 34, n. 2, p. 253-270, 1928.

Sobre o autor

Jerônimo Oliveira Muniz é Ph.D. em Sociologia/ Demografia pela University of Wisconsin, Madison. Professor e pesquisador do Programa de Pós-graduação em Sociologia da Universidade Federal de Minas Gerais (UFMG).

Endereço para correspondência

Faculdade de Filosofia e Ciências Humanas (Fafich)
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627, Fafich 4156, *campus* Pampulha
31270-901 – Belo Horizonte-MG, Brasil

Abstract

Demographic analysis and forecasting using of inter-municipal population growth and distribution matrices

Challenges in the field of demographic projections include, among others, the volatility of the migration component - critical for the projection of small areas; the compatibility between projections of small and large areas; and the measurement and inclusion of uncertainty in future scenarios of population growth. This article presents a new probabilistic method to conduct interregional population forecasting dealing with these three challenges. The proposed method has the following advantages: 1) it only requires information about the last place of residence and the population distributions of the last two Censuses; 2) it generates confidence intervals

for the projected populations; 3) it makes the role of migration flows in the growth dynamics explicit and; 4) it facilitates the elaboration of counterfactual scenarios and sensitivity analysis using matrices of interregional population growth and distribution. We describe the patterns and trends in migration flows in the state of São Paulo applying spatial visualization tools and identifying areas in which migration is responsible for considerable shares of the demographic dynamics. About 95% of the 572 municipal projected populations of São Paulo had good precision and were within expected confidence intervals. We used data from the 1980, 1991 and 2000 Brazilian Censuses.

Keywords: Projections. Migration. County. Natural growth. Spatial analysis.

Resumen

Análisis y previsión demográfica utilizando matrices de crecimiento y distribución intermunicipal de la población

Entre los desafíos en el campo de las proyecciones demográficas están la volatilidad del componente de migración - fundamental para la proyección de áreas pequeñas; la compatibilidad entre las proyecciones de las áreas grandes y pequeñas; y la medición y la inclusión de la incertidumbre en los escenarios futuros de crecimiento de la población. Este artículo presenta un nuevo método probabilístico para llevar a cabo las proyecciones de población interregional que se ocupa de estos tres desafíos. Las ventajas del método propuesto son: utilizar como entrada únicamente la información sobre el municipio de residencia y las distribuciones de población de los dos últimos censos; generar intervalos de confianza para la población proyectada; enfatizar el papel de la migración en la dinámica de crecimiento inter-municipal; y facilitar el desarrollo de escenarios hipotéticos y análisis de sensibilidad utilizando matrices de crecimiento multirregional. Describe los patrones y tendencias de los flujos migratorios en el estado de São Paulo aplicando herramientas de visualización espacial y identifica las áreas en que la migración se encarga de parte considerable de la dinámica demográfica observada. Más del 95% de las 572 poblaciones proyectadas de los municipios de São Paulo tuvieron buena precisión y se quedaron dentro de los intervalos de confianza especificados. Utilizamos los datos de los Censos de Brasil de 1980, 1991 y 2000.

Palabras clave: Proyecciones. Migración. Municipio. Crecimiento natural. Análisis espacial.

Recebido para publicação em 25/04/2018

Aceito para publicação em 18/07/2018

Anexo 1

Programa de Stata utilizado para estimar os coeficientes bayesianos, realizar as previsões populacionais a partir da matriz de crescimento proposta e replicar os gráficos apresentados nos Anexos 2 e 5

```
*****
*PROGRAMA PARA ESTIMACAO DE COEFICIENTES BAYESIANOS DA MATRIZ DE CRESCIMENTO*
*****
set more off
clear all

mata mata matuse matriz91.mmat, replace
mata mata matuse matriz00.mmat, replace
mata vec91=vec(G1trans91)
mata vec00=vec(G1trans00)
set obs 328329 /*creates 328.329 empty rows in the dataset*/
mata st_store(.,st_addvar("float", "vec91"),vec91)
mata st_store(.,st_addvar("float", "vec00"),vec00)

*****
*Gamma likelihood function: link identity *
*Evaluator kindly provided by Prof. Hilbe in the end of 2015 and later *
*published in: Hardin, J. and Hilbe, J. 2018. Generalized Linear Models and *
*Extensions, Fourth Edition. A Stata Press Publication StataCorp LLC College *
*Station, Texas. *
*****
program define bgammall
    version 15
    args lnf xb phi
    quietly {
        tempvar lnfj
        generate double `lnfj'= (1/`phi') * log($MH_y/ (`phi'*`xb')) - ///
($MH_y/ (`phi'*`xb')) -log($MH_y)- lngamma(1/`phi') if $MH_touse
        summarize `lnfj' if $MH_touse, meanonly
        if r(N) < $MH_n {
            scalar `lnf' = .
            exit
        }
        scalar `lnf' = r(sum)
    }
end
*Bayesian estimates using gamma likelihood function and uninformative priors
bayesmh vec00 vec91 if vec00>0, llevaluator(bgammall, parameters({phi})) ///
prior({vec00:} {phi}, flat) ///
block({phi}) init({vec00:_cons} 1 {phi} 1) ///
rseed(666) saving(b1, replace) mcmcsz(50000)est store b1

*Diagnostics
bayesgraph diagnostics _all
bayesgraph cusum _all
bayesstats ic b*
bayesstats ess _all
bayesgraph matrix {vec91}{_cons}, half

/*Frequentist estimates: for comparison
bootstrap, seed(666) reps(100): glm vec00 vec91, fam(gamma) link(id) nolog
predictnl mu = predict(), ci(mu_left mu_right) level(95) /*95%CI and predicted
values*/
scatter vec00 vec91 || line mu vec91 || line mu_left vec91 || line mu_right
vec91
*/
```

```

*****
*CODE FOR THE PREDICTION OF THE MAIN DIAGONAL USING BAYESIAN REGRESSION*
*****
use "module4.dta", clear
*Frequentist estimates:
qui reg diagG00 diagG
predict rstudent if e(sample), rstudent
rreg diagG00 diagG, genwt(peso) nolog /*Robust regression gives lower weight to
outliers*/
predictnl mi = predict(), ci(mi_left mi_right) level(95)
scatter diagG00 diagG if peso>0 || rarea mi_left mi_right diagG ///
|| scatter diagG00 diagG if peso==0, mlabel(rstudent) || line mi diagG

*Bayesian estimates: OLS without influent outliers and uninformative priors
bayesmh diagG00 diagG if peso>0, likelihood(normal({var})) ///
prior({_cons}, normal(0,1000)) ///
prior({diagG}, normal(0,1000)) ///
prior({var}, igamma(0.1,0.1)) ///
block({diagG00: _cons}, gibbs) ///
block({diagG00: diagG}, gibbs) ///
block({var}, gibbs) ///
rseed(666) mcmcsize(50000) saving(b2, replace) thinning(4)

est store b2
*Diagnostics
bayesgraph diagnostics _all
bayesgraph cusum _all
bayesstats ic b*
bayesstats ess _all
bayesgraph matrix {diagG}{_cons}, half

*****
*PROGRAMA PARA PREVISAO E PROJECAO DEMOGRAFICA DOS MUNICIPIOS PAULISTAS,v.1.0*
*Desenvolvido na versão 14.1 do Stata *
*****
set more off
clear all
program define previpop, rclass
drop _all
*****
*GENERATES 1000 PROJECTION VECTORS for the off-diagonal*
*****
use pij1991.dta, clear
qui su pij1991
forvalues i = 1(1)1000 {
generate x`i' = pij1991*rnormal(1,r(sd))
}
*****
*PRODUCES 1000 MIGRATION MATRICES OFF-DIAGONAL*
*****
mata mata matuse matriz91.mmat, replace
mata st_view(TEMP=.,.,) /*creates a matrix from the data set*/
forvalues k=1/1000 {
mata mat`k'=I(573)
mata st_subview(mat`k',TEMP,(1,328329),`k'+1)
mata mat`k'_91=mat`k'
mata mat`k'_91=rowshape(mat`k'_91',573)'
mata mata drop mat`k'
}

```

```

*****
*GENERATES 1000 PROJECTION VECTORS for the DIAGONAL*
*****
use ci1991.dta, clear
qui su ci1991
forvalues i = 1(1)1000 {
    generate x`i' = ci1991*rnormal(1,r(sd))
}
*****
*PRODUCES 1000 NATURAL GROWTH MATRICES ON DIAGONAL*
*****
mata st_view(TEMPd=.,.,.)
forvalues k=1/1000 {
    mata matd`k'=I(1)
    mata st_subview(matd`k',TEMPd,(1,573),`k'+1)
    mata matd`k'_91=diag(matd`k')
    mata mata drop matd`k'
}
*****
*1000 PROJECTIONS with simulated matrices*
*****
forvalues k=1/1000 {
    mata pop`k'=(mat`k'_91+matd`k'_91)'*(mat`k'_91+matd`k'_91)'+p90
}
*****
*Adds the current, observed and projected populations to the dataset*
*****
forvalues k=1/1000 {
    mata st_store(.,st_addvar("float","pop`k'"),pop`k')
}
mata st_store(.,st_addvar("float","p90"),p90)
mata st_store(.,st_addvar("float","p00"),p00)
*****
*Counts the number of intervalar municipal projections*
*for 2000 including the observed population *
*****
    keep pop1-pop1000 p90 p00
    egen rowmax=rowmax(pop1-pop1000)
    egen rowmin=rowmin(pop1-pop1000)
    egen rowmean=rowmean(pop1-pop1000)
    gen count=cond(p00>=rowmin & p00<=rowmax,1,0)
    quiet su count
    return scalar mean= r(mean)
    return scalar sd=r(sd)
    mata mata clear
end
*****
/*Performs 100 Monte Carlo Simulations: very time consuming*
*****
simulate mean=r(mean) sd=r(sd), rep(100): previpop
su /*mean represents the percent of municipalities (out of 573) whose interval
    includes the real observed population*/
*/

```

```

*****
*Precision measures for the forecasting: based on 1000 projections*
*****
use inputAiBi.dta, clear
gen EMP=((rowmean-p00)/p00)*100 in 1/572
quiet su EMP in 1/572, d
di r(sum)/r(N) /*EMP*/
gen absEMP=abs(EMP)
quiet su absEMP in 1/572, d
di r(sum)/r(N) /*EMPM*/
gen ratio=p00/rowmean
quiet su rowmean in 1/572, d
gen ratio2=p00/(rowmean*(36102875/r(sum))) /*adjusts the mean forecast for
the projection of the state population*/
su ratio ratio2 in 1/572 /*ROP*/

*****
*Projections based on AiBi method, for comparison. The method has three *inputs:
municipal and state pop. of 1980 and 1991, and proj. state pop. for *2000
*****
egen tot91=total(p1991) in 1/572
egen tot80=total(p1980) in 1/572
gen ai=(p1991-p1980)/(tot91-tot80) in 1/572
gen bi= p1991-(ai*tot91) in 1/572
gen Proj00=ai*36102875+bi in 1/572
label var Proj00 "AiBi 2000 projections"
/*36.102.875 represents projected population for the State of Sao Paulo in 2000,
published in: "Waldvogel and Capassi (1999). Cenarios da populacao paulista: dos
anos 90 ao futuro. Sao Paulo em perspectiva 13(1-2)" */
*****
*Precision measures AiBi*
*****
gen EMPAiBi= ((Proj00-p00)/p00)*100 in 1/572
quiet su EMPAiBi in 1/572, d
di r(sum)/r(N) /*EMP: erro medio percentual da projecao AiBi*/
gen absEMPAi=abs(EMPAiBi)
quiet su absEMPAi in 1/572, d
di r(sum)/r(N) /*EMPM: EMP em módulo*/
gen ratioAiBi= p00/Proj00 in 1/572
su ratioAiBi in 1/572 /*ROP: razão entre pops. projetadas e observadas*/

*****
*Precision measures for the deterministic projection*
*****
mata mata matuse matriz91.mmat
mata proj00=round(G91*G91*p90) /*2000 deterministic population projection*/
mata st_store(.,st_addvar("float","proj00"),proj00)
label var proj00 "2000 deterministic projections"
gen EMPdet= ((proj00-p00)/p00)*100 in 1/572
quiet su EMPdet in 1/572, d
di r(sum)/r(N) /*EMP: erro medio percentual da projecao deterministica*/
gen absEMPdet=abs(EMPdet)
quiet su absEMPdet in 1/572, d
di r(sum)/r(N) /*EMPM: EMP em módulo*/
gen ratioDET=p00/proj00 in 1/572
su ratioDET in 1/572 /*ROP: razão entre pops. projetadas e observadas*/

*COMPARATIVE HISTOGRAMS of the error*
two hist EMPAiBi if EMPAiBi<40, freq || hist EMP if EMP<40, freq || ///
hist EMPdet if EMPdet<40, freq legend(on all)

*****

```

```

*ANEXO 1*
*****
use module4.dta, clear
*Distribuição e correlação entre elementos pij, fora da diagonal principal
scatter vec00 vec91 if vec00!=0 ///
    || lfitci vec00 vec91 if vec00!=0, yscale(alt) xscale(alt) ///
    xlabel(, grid gmax) saving(yx, replace) legend(off) scheme(lean1)
hist vec00 if vec00!=0, fraction xscale(alt reverse) horiz fysize(25) ///
    saving(hy, replace) scheme(lean1)
hist vec91 if vec00!=0, fraction yscale(alt reverse) ylabel(0(.2).8, nogrid) ///
    xlabel(,nogrid) fysize(25) saving(hx, replace) scheme(lean1)
gr combine hy.gph yx.gph hx.gph, hole(3) ///
    imargin(0 0 0 0) graphregion(margin(l=22 r=22)) scheme(lean1)

*Distribuição e correlação entre elementos ci, sob a diagonal principal
scatter diagG00 diagG ///
    || lfitci diagG00 diagG, yscale(alt) xscale(alt) ///
    xlabel(, grid gmax) saving(yx, replace) legend(off) scheme(lean1)
hist diagG00, percent xscale(alt reverse) horiz fysize(25) ///
    saving(hy, replace) scheme(lean1)
hist diagG, percent yscale(alt reverse) ylabel(0(.2).8, nogrid) ///
    xlabel(,grid gmax) fysize(25) saving(hx, replace) scheme(lean1)
gr combine hy.gph yx.gph hx.gph, hole(3) ///
    imargin(0 0 0 0) graphregion(margin(l=22 r=22)) scheme(lean1)

*****
*ANEXO 4*
*****
use 1000_projections.dta, clear
xpose, clear varname format promote
set more off

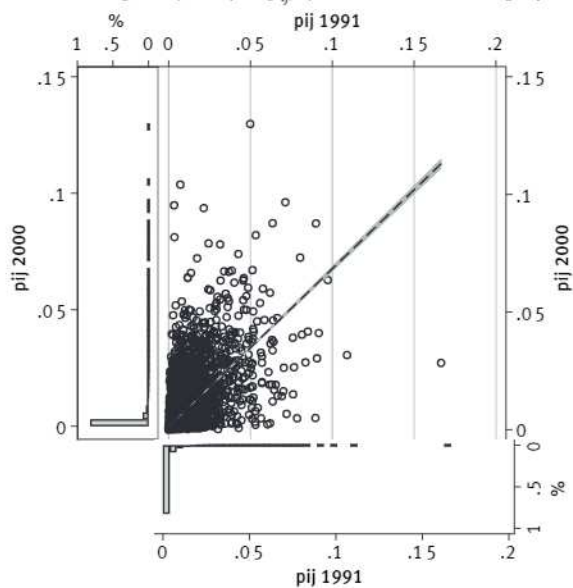
forvalues i=1(1)573 {
hist v`i' in 1/1001, frequency lcolor(white) lpattern(blank) normal ///
    normopts(lpattern(dash)) kdensity addplot((dropline v`i' v`i' if v`i' in 1001, ///
    yaxis (2) mcolor(red) msize(large) msymbol(circle) mlabel(v`i') legend(off) ///
    mlabsize(default) mlabposition(12) lcolor(red) lwidth(thick) lpattern(solid)) ///
    ytitle(Frequência) xtitle(População) scheme(lean1) saving(v`i'.gph, replace)
}
gr combine v77.gph v149.gph v225.gph v518.gph v349.gph v26.gph v389.gph ///
    v365.gph v210.gph v362.gph v407.gph v486.gph, colfirst ycommon imargin(medlarge) ///
    scheme(lean1)

```

Anexo 2

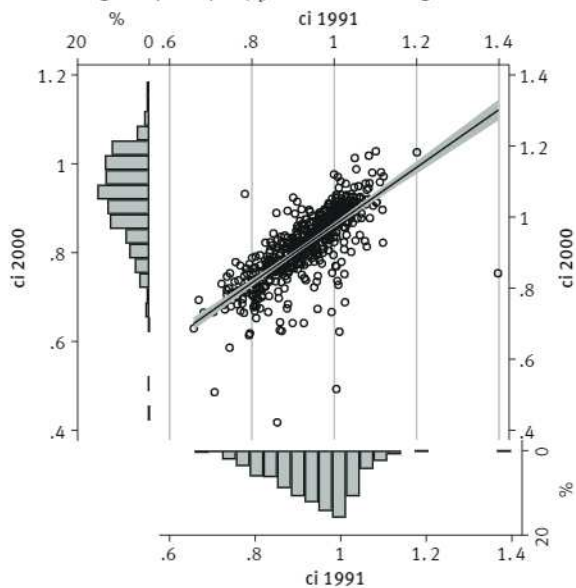
Distribuição e correlação entre os elementos da matriz de crescimento

Elementos fora da diagonal principal (p_{ij}): probabilidades de migração de i para j



Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Elementos da diagonal principal (c_i): crescimento vegetativo mais emigração



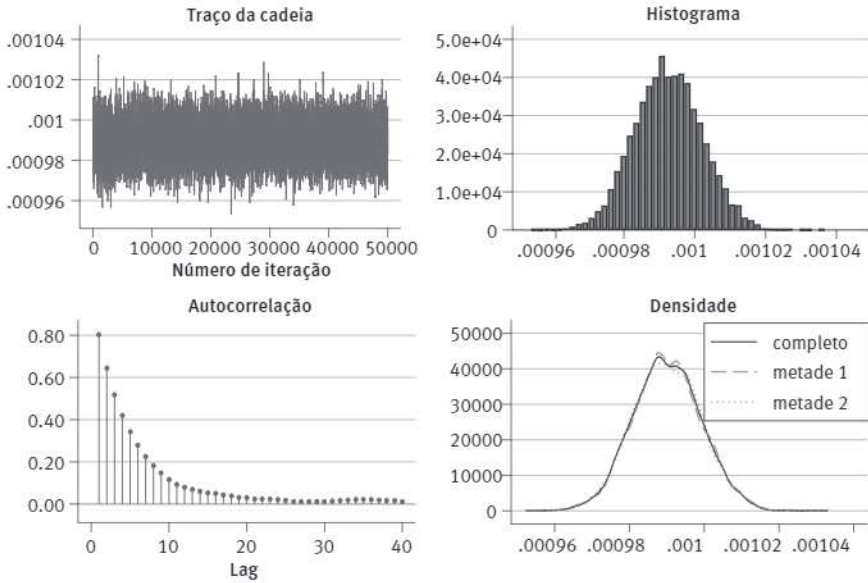
Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Anexo 2.1

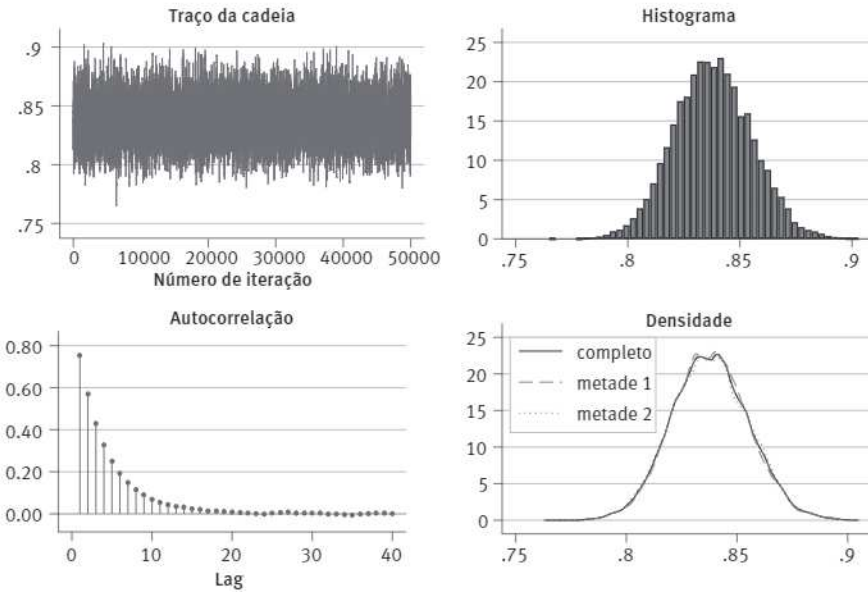
Medidas diagnósticas de convergência dos coeficientes bayesianos estimados

Modelo linear generalizado da família gama, com função de ligação identidade, para a estimação de coeficientes correlacionando probabilidades de migração intermunicipais
Estado de São Paulo – 1991-2000

B0



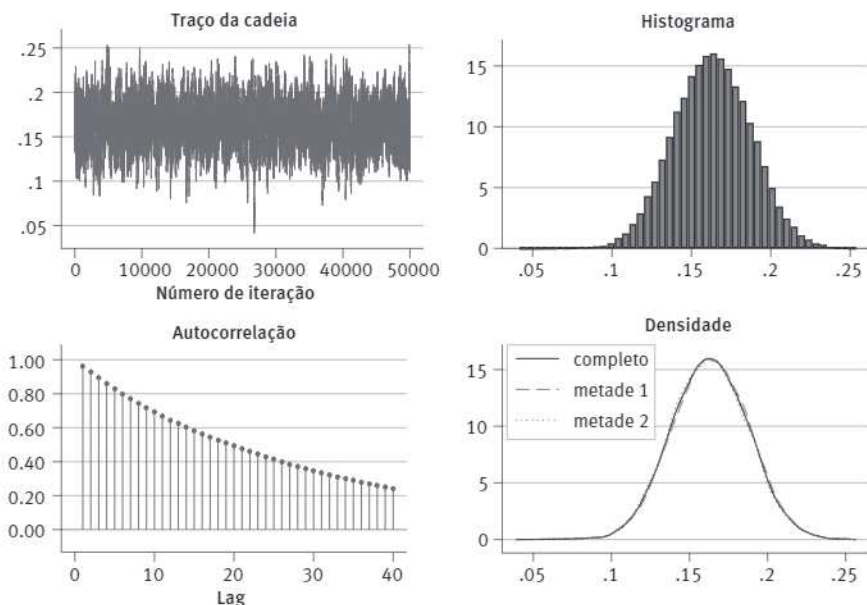
B1 (pij 1991)



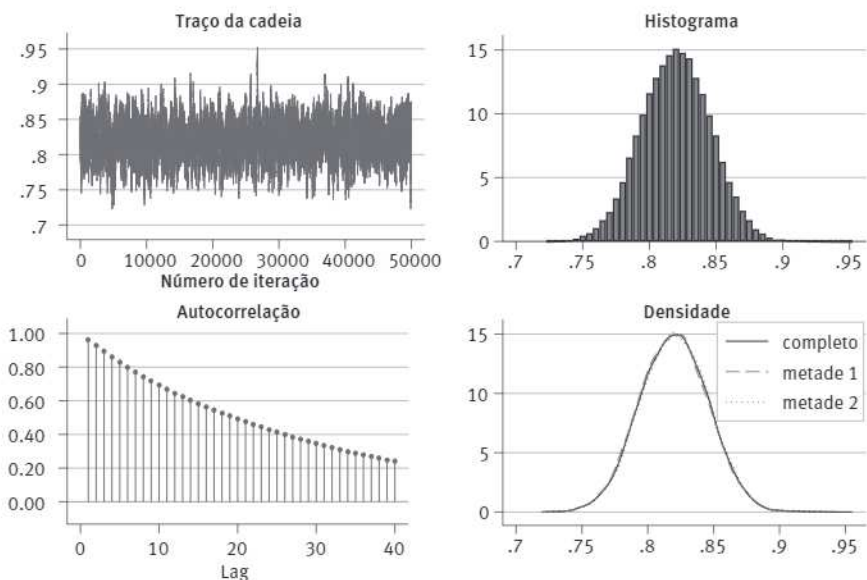
Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Modelo bayesiano com função de máxima verossimilhança linear normal, para a estimação de coeficientes correlacionando o componente correspondente ao crescimento natural e emigração intermunicipal da matriz de crescimento Estado de São Paulo – 1991-2000

a0



a1 (ci 1991)



Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Anexo 3

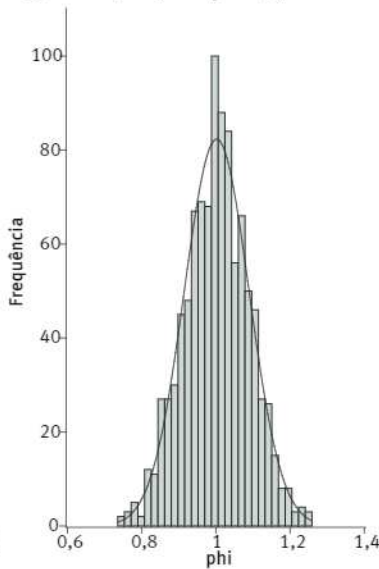
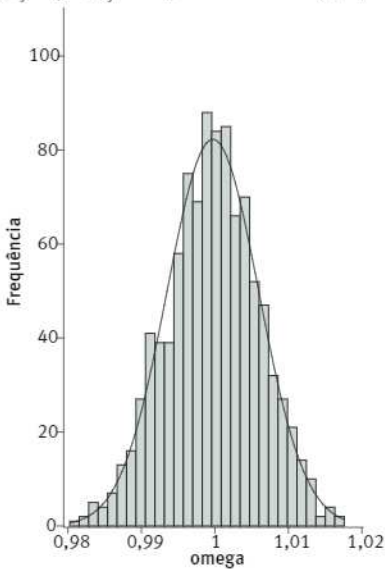
Parâmetros de variação da matriz de crescimento

```

*****
*Programa de Stata para gerar 1000 parâmetros de flutuação para as *
matrizes de crescimento e seus respectivos histogramas de frequência *
*****
matrix m = (1,1)
matrix sd = (.0063,.0909) /*define os desvios-padrão empíricos*/
drawnorm omega phi, n(1000) means(m) sds(sd)
summarize
hist omega, freq norm saving(omega_his.gph, replace)
hist phi, freq norm saving(phi_his.gph, replace)
gr combine omega_his.gph phi_his.gph, ycommon
    
```

$$p_{ij}^{1991}_u = p_{ij}^{1991} \varpi, \text{ sendo } \varpi \sim N(1; 0,0063)$$

$$c_i^{1991}_u = c_i^{1991} \phi, \text{ sendo } \phi \sim N(1; 0,0909)$$

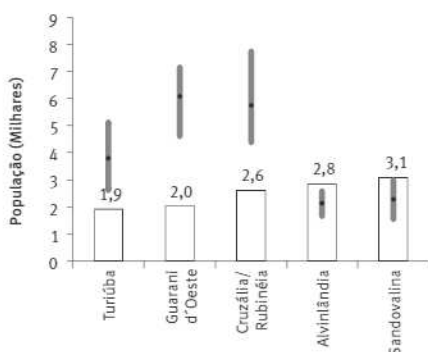


Parâmetro	Obs	Média	Des. Padrão	Min	Max
omega	1000	.9998105	.0060189	.9805955	1.02022
phi	1000	.9988998	.0901226	.7123487	1.304946

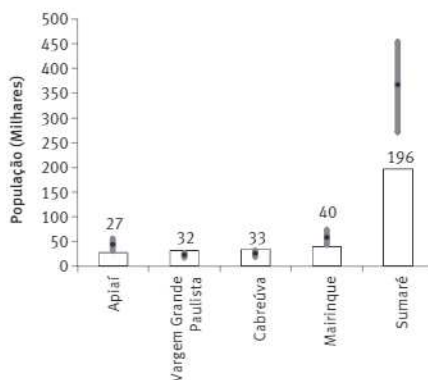
Anexo 4

Previsões populacionais municipais selecionadas

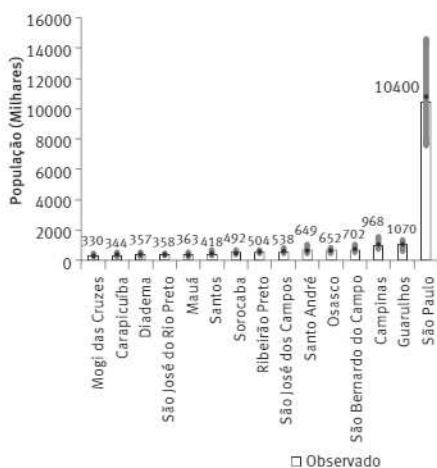
Cinco **menores** municípios com populações observadas **fora** do intervalo populacional probabilístico



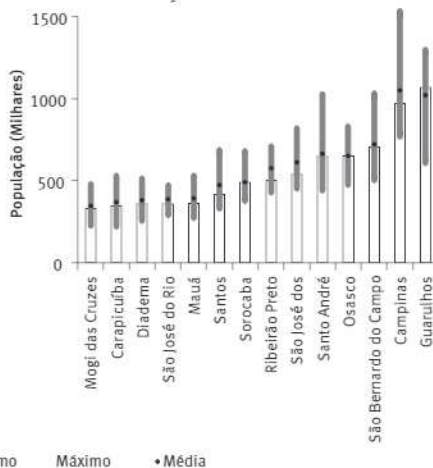
Cinco **maiores** municípios com populações observadas **fora** do intervalo populacional probabilístico



Quinze **maiores** municípios com populações observadas **dentro** do intervalo populacional probabilístico

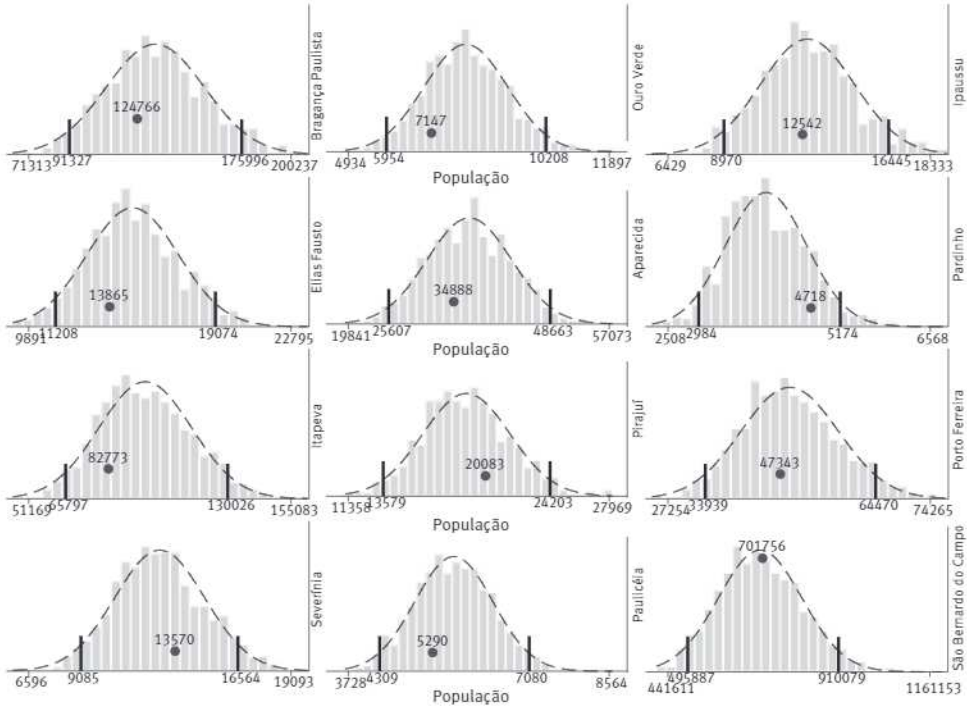


Catorze **maiores** municípios com populações observadas **dentro** do intervalo populacional probabilístico (excluindo São Paulo)



Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Anexo 5

Previsões populacionais *ex-post* para 12 municípios paulistas aleatoriamente selecionados

Fonte: IBGE. Censos Demográficos de 1991 e 2000.

Nota: As linhas verticais pretas demarcam os intervalos de 95% de confiança para as populações municipais projetadas. Os valores mínimos e máximos projetados para cada município são apresentados no eixo horizontal. A população observada em 2000 é representada pelos pontos no interior da distribuição de frequência das populações projetadas. O eixo vertical varia entre 0 e 800 mil. A flutuação vertical dos pontos representando as populações municipais observadas deve ser ignorada.