

CONSTRUÇÃO DA MATRIZ ORIGEM-DESTINO DE TRANSPORTE INTER-REGIONAL DE CARGAS E PASSAGEIROS PARA O PLANO NACIONAL DE LOGÍSTICA INTEGRADA

Relatório de Pesquisa

Métodos de Estimação da Matriz Origem-destino para o Transporte Aéreo de Passageiros (Relatório 11)



CONSTRUÇÃO DA MATRIZ ORIGEM-DESTINO DE TRANSPORTE INTER-REGIONAL DE CARGAS E PASSAGEIROS PARA O PLANO NACIONAL DE LOGÍSTICA INTEGRADA

Relatório de Pesquisa

Métodos de Estimação da Matriz Origem-destino para o Transporte Aéreo de Passageiros (Relatório 11)

ipea

Governo Federal

Ministério do Planejamento, Desenvolvimento e Gestão
Ministro interino Dyogo Henrique de Oliveira

ipea Instituto de Pesquisa Econômica Aplicada

Fundação pública vinculada ao Ministério do Planejamento, Desenvolvimento e Gestão, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente

Ernesto Lozardo

Diretor de Desenvolvimento Institucional, Substituto

Carlos Roberto Paiva da Silva

Diretor de Estudos e Políticas do Estado, das Instituições e da Democracia

Alexandre de Ávila Gomide

Diretor de Estudos e Políticas Macroeconômicas

José Ronaldo de Castro Souza Júnior

Diretor de Estudos e Políticas Regionais, Urbanas e Ambientais

Alexandre Xavier Ywata de Carvalho

Diretor de Estudos e Políticas Setoriais de Inovação e Infraestrutura

João Alberto De Negri

Diretora de Estudos e Políticas Sociais

Lenita Maria Turchi

Diretor de Estudos e Relações Econômicas e Políticas Internacionais

Sérgio Augusto de Abreu e Lima Florêncio Sobrinho

Assessora-chefe de Imprensa e Comunicação

Regina Alvarez

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

URL: <http://www.ipea.gov.br>

CONSTRUÇÃO DA MATRIZ ORIGEM-DESTINO DE TRANSPORTE INTER-REGIONAL DE CARGAS E PASSAGEIROS PARA O PLANO NACIONAL DE LOGÍSTICA INTEGRADA

Relatório de Pesquisa

Métodos de Estimação da Matriz Origem-destino para o Transporte Aéreo de Passageiros (Relatório 11)

ipea

EQUIPE DE PESQUISA

Coordenação-geral:

Fabiano Mezadre Pompermayer, da Diretoria de Estudos e Políticas Setoriais de Inovação, Regulação e Infraestrutura (Diset) do Ipea

Equipe técnica

Fabiano Mezadre Pompermayer, da Diset do Ipea

Erivelton Pires Guedes, da Assessoria Técnica da Presidência (Astep) do Ipea

Akina Sakamori, bolsista do Programa de Pesquisa para o Desenvolvimento Nacional (PNPD) da Diset do Ipea

Alan Ricardo da Silva, bolsista do PNPD da Diset do Ipea

Carolina Andrade Silva, bolsista do PNPD da Diset do Ipea

Daniel Alisson Feitosa Lopes, bolsista do PNPD da Diset do Ipea

Diego Rosa Mambrin, bolsista do PNPD da Diset do Ipea

Gabriel Gouveia Rabello, bolsista do PNPD da Diset do Ipea

João Gabriel de Moraes Souza, bolsista do PNPD da Diset do Ipea

Maircon Batista Ribeiro, bolsista do PNPD da Diset do Ipea

Paulo Henrique Dourado da Silva, bolsista do PNPD da Diset do Ipea

Pedro Veiga de Camargo, bolsista do PNPD da Diset do Ipea

Priscila Nascimento de Alcântara Garcia, bolsista do PNPD da Diset do Ipea

Rafaella Bandeira Cabral Cunha, bolsista do PNPD da Diset do Ipea

Ramon de Almeida Bispo, bolsista do PNPD da Diset do Ipea

Raquel Araujo de Almeida, bolsista do PNPD da Diset do Ipea

Rennaly Patricio Sousa, bolsista do PNPD da Diset do Ipea

Roberto Lazarte Kaqui, bolsista do PNPD da Diset do Ipea

Thiago Guimarães Rodrigues, bolsista do PNPD da Diset do Ipea

Equipe de elaboração do relatório:

Thiago Guimarães Rodrigues, bolsista do PNPD da Diset do Ipea

Fabiano Mezadre Pompermayer, da Diset do Ipea

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério do Planejamento, Desenvolvimento e Gestão.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

SUMÁRIO

APRESENTAÇÃO	8
1 INTRODUÇÃO	9
2 MÉTODOS DE ESTIMATIVA BASEADOS EM DADOS DE CONTAGEM	10
3 RECOMENDAÇÕES DE MODELAGEM	13
4 INÍCIO DE APLICAÇÃO DA PROPOSTA: DESENVOLVIMENTO DO MODELO DE GERAÇÃO DE VIAGENS AÉREAS DE PASSAGEIROS	18
5 CONSIDERAÇÕES FINAIS	26
REFERÊNCIAS	27

APRESENTAÇÃO

Este relatório faz parte de uma série de relatórios de pesquisa do Ipea, cujo objetivo é documentar o processo de construção das matrizes origem/destino (O/D) de transporte inter-regional de cargas e passageiros que subsidiaram a elaboração do Plano Nacional de Logística Integrada (PNLI), fruto da parceria entre a Empresa de Planejamento e Logística (EPL), empresa pública do governo federal, e o Ipea, formalizada pelo Termo de Cooperação para Descentralização de Crédito no 2, de 29 de maio de 2013.

A EPL foi criada em 2012 para apoiar o planejamento da infraestrutura de transportes do país. Uma de suas tarefas é elaborar o Plano Nacional de Logística Integrada. Em certa medida, ela cumpre missão semelhante à do Grupo de Estudos para Integração da Política de Transportes (Geipot) nas décadas de 1970 e 1980. A empresa vem se capacitando para a modelagem da oferta de infraestrutura de transportes, mas para a modelagem da demanda por transportes, que envolve mais conhecimento sobre as atividades econômicas no país e sua distribuição regional do que sobre transportes, havia a necessidade de execução externa. Algumas consultorias foram abordadas, mas surgiu a possibilidade de o estudo ser realizado pelo Ipea, que prontamente atendeu.

Nesse processo, um dos passos iniciais é o levantamento do padrão de viagens inter-regionais de cargas e passageiros, identificando os potenciais de geração e atração de viagens de cada localidade, bem como sua distribuição entre elas. Os modos de transporte usados também são levantados. Assim, se obtém a matriz origem-destino das viagens de carga e passageiro para um ano-base, em cada modo. A partir daí, são projetados os crescimentos de demanda por transporte para os anos futuros – isto é, as matrizes O/D futuras, que são confrontadas com a oferta de infraestrutura, a fim de identificar seus gargalos e projetar a necessidade de melhorias, como construção ou ampliação de rodovias, ferrovias, hidrovias e portos. A construção da matriz O/D é essencial para o planejamento da infraestrutura de transportes do país, uma vez que permite a construção de cenários e a otimização dos investimentos.

A elaboração de uma matriz O/D envolve o levantamento e o cruzamento de uma série de dados socioeconômicos, que são ajustados e calibrados com pesquisas de campo, em que viajantes são entrevistados em locais predeterminados para identificar diversos atributos de sua viagem, como origem, destino, motivo, tipo e valor da carga, modo(s) de transporte utilizado(s) etc. Esse trabalho envolve conhecimento técnico específico em modelagem de transportes e também em economia e estatística.

Com uma matriz O/D consolidada e abrangente, a EPL poderá realizar suas avaliações de necessidade de expansão da oferta de infraestrutura de transporte. O Ipea, além de apoiar a EPL nestas avaliações, se beneficia das informações da matriz O/D, ao cruzá-las com outras fontes de dados socioeconômicos, podendo realizar estudos diversos sobre o desenvolvimento regional, as matrizes insumo-produto e o fomento produtivo, por exemplo. Adicionalmente, a disponibilização dessa base de dados à sociedade permite o desenvolvimento do próprio setor de transporte e logística.

Oito relatórios descrevem as metodologias usadas em todo o processo de construção da base de dados e suas projeções. O último desses relatórios faz a agregação dos procedimentos intermediários para a construção das matrizes e apresenta alguns dos resultados agregados. Há, ainda, três relatórios que avaliaram o estado da arte de modelos de construção da matriz O/D, com proposições para as novas versões do PNLI. A seguir, a lista com os títulos de cada relatório.

1. *Desenho da Pesquisa Origem-Destino do Transporte Rodoviário no Brasil.*
2. *Construção da Matriz Origem-Destino Observada para Transporte de Cargas Inter-regional.*
3. *Modelos de Regressão para Geração e Atração de Viagens do Transporte de Cargas Inter-regional.*
4. *Modelos de Distribuição para Matriz Origem-Destino de Transporte de Cargas Inter-Regional: desenvolvimento de um conjunto de ferramentas e calibração inicial.*
5. *Cenários de Projeção das Atividades Econômicas por UF para a Projeção da Matriz Origem-Destino de Transporte de Cargas Inter-regional.*
6. *Construção da Matriz Origem-Destino Observada de Transporte de Passageiros Inter-regional.*
7. *Modelos de Regressão para Geração e Atração de Viagens de Passageiros.*
8. *Projeções das Matrizes Origem-Destino de Carga e Passageiros: aplicando os modelos de geração e distribuição de viagens com as projeções socioeconômicas.*
9. *Estado da Arte em Métodos de Construção de Matrizes Origem-Destino para o Transporte de Cargas Inter-regional.*
10. *Estado da Arte de Métodos de Estimação de Matrizes Origem-Destino para Passageiros a Longa Distância.*
11. *Métodos de Estimação da Matriz Origem-Destino para o Transporte Aéreo de Passageiros.*

1 INTRODUÇÃO

Um passo crucial da modelagem de transporte consiste em estimar matrizes origem-destino precisas, que contenham informações sobre os volumes de tráfego entre cada par de zonas em uma área de estudo. Vários métodos de estimação podem ser utilizados, com a finalidade de determinar os fluxos (valores de células de uma matriz O/D) que se encaixam bem para valores empiricamente avaliados; por exemplo, aqueles avaliados por contagens de tráfego ou pesquisas em um ano-base. A capacidade de se replicar fluxos observados de tráfego é vista como um dos requisitos mais importantes do modelo (FHWA, 2013).

A maioria das aplicações conta com o paradigma do modelo de transporte de demanda de quatro etapas, que calcula na etapa de geração o agregado de demanda de viagens, baseado em informações socioeconômicas, demográficas e de uso do solo, no nível de zona. No entanto, o desenvolvimento de tais modelos requer uma grande quantidade de dados levantados por pesquisas domésticas, que são normalmente muito caras e demoradas. Além disso, os dados recolhidos por tais pesquisas podem ser rapidamente ultrapassados em contextos de mudanças – como nos países em desenvolvimento.

Uma estrutura do modelo mais simples corresponde ao uso de dados observados para a estimativa da matriz origem-destino, a partir da – e para inferir a – alocação de tráfego. O principal tipo de dados de entrada para a estimação do modelo é um conjunto de contagens de tráfego obtido por meio do monitoramento da operação de serviço. Como a demanda é derivada de contagens de tráfego, essa abordagem carece de detalhes comportamentais e não permite a análise de sensibilidade apenas utilizando valores das variáveis demográficas ou econômicas. Essa abordagem alternativa permite modelar a escolha de diferentes vias do mesmo par origem-destino, mas não leva em conta a concorrência entre os diferentes modos.

A estimação do modelo de contagens de tráfego é cada vez mais utilizada, mas ainda raramente aplicado na prática de modelagem atual. Há também uma falta de relatórios detalhados sobre a sua aplicação em modelos operacionais. O interesse deste trabalho é o método de estimativa de fluxo de tráfego de modo específico de partida, a partir de dados empíricos disponíveis (os problemas relacionados com a própria avaliação de dados – ou seja, determinar o melhor conjunto de locais de contagem de tráfego em uma determinada rede, não são objeto deste estudo). O objetivo é apresentar o estado da arte em técnicas de estimativa e produzir recomendações concretas, para estimar uma matriz O/D baseada em contagens de transporte aéreo de passageiros para cada *link* de tráfego aéreo no Brasil.

Este trabalho tem como objetivo desenvolver recomendações metodológicas de um procedimento de estimação da matriz O/D, com base no volume de tráfego observado entre aeroportos em nível nacional. Primeiramente, apresentamos, na seção 3, as principais abordagens de estimação da matriz O/D e as questões críticas relacionadas. A seção 4 concentra-se em recomendações concretas para a concepção de um modelo baseado em dados disponíveis. A seção 5 faz um exercício com modelo de geração de viagens para exemplificar parte da abordagem proposta na seção 4. A última seção apresenta as considerações finais.

2 MÉTODOS DE ESTIMATIVA BASEADOS EM DADOS DE CONTAGEM

A estimação de uma matriz O/D a partir de dados empíricos – por exemplo, contagens de tráfego – é considerada por Bera e Rao (2011) como “um dos temas mais promissores para pesquisa”. A tarefa, por vezes referida como o “problema de alocação inversa”, consiste na obtenção de matrizes O/D a partir dos fluxos observados em *links* (arcos da rede de transportes) selecionados conectando diferentes zonas. Contagens de volumes de tráfego normalmente fornecem uma matriz origem-destino parcial, que pode ser então expandida para cobrir toda a área de estudo. A sua atratividade está ligada à relativa facilidade de se obter esses dados de uma forma não disruptiva e aos custos mais baixos, em comparação com os métodos tradicionais.

De acordo com Abrahamsson (1998), o problema de estimação é encontrar uma matriz O/D razoável que, quando atribuída à rede de transporte, reproduz os dados de contagem de tráfego observados. Entretanto, uma reprodução exata para cada contagem de dados pode não ser adequada, devido a questões técnicas de diferentes tempos de coleta.

Formalmente, a estimação da matriz O/D pela alocação inversa é um problema matemático com várias soluções possíveis, uma vez que o número de contagens não é, geralmente, suficiente para determinar todos os fluxos por par O/D. Embora essa questão tenha sido intensamente explorada nas últimas décadas – em particular no final dos anos 1980 –, não parece haver consenso sobre quais das várias abordagens propostas, com base em vários algoritmos e procedimentos de estimativa, devem ser preferidas para gerar uma solução plausível. Esta seção se baseia, principalmente, nas avaliações de Abrahamsson (1998) e Bera e Rao (2011), que fornecem uma revisão sistemática desses métodos.

Uma diferença importante entre os métodos de estimação da matriz O/D é a sua capacidade de levar em conta os tempos de viagem em condições reais. Bera e Rao (2011) referem-se a *modelos de estimação estáticas e dinâmicas*.

- 1) *Modelos estáticos* são os em que fluxos são considerados independentemente das variações do tempo de viagem – como efeitos de congestionamentos – e a matriz resultante representa uma situação de estado estacionário. Como consequência, a matriz contendo a proporção de viagens entre um par de zonas usando determinados *links* pode ser estimada pelo procedimento de alocação proporcional. Por esse procedimento, proporções de viagens por *link* não dependem do nível de congestionamento de cada *link*, mas apenas das características de viajantes e rotas.
- 2) *Modelos de estimação dinâmicos* – ou técnicas de alocação não proporcionais – consideram os efeitos do congestionamento. Métodos de alocação de equilíbrio são usados para determinar o custo da viagem em um *link* como uma função do volume de fluxo (função de restrição de capacidade). Técnicas de alocação de equilíbrio seguem o primeiro princípio de equilíbrio de Wardrop. Esses métodos fornecem matrizes mais realistas em redes em que os tempos de viagem reais diferem dos tempos de viagem de fluxo livre, mas são matematicamente mais complexos, porque eles exigem uma estimativa iterativa de escolha de rotas e matrizes de viagem. Assim, essa técnica tem sido utilizada principalmente para estudos de pequena escala. Conforme Bera e Rao (2011, tradução nossa), “em comparação com a estimativa de matriz O/D estática, estudos baseados em estimativa dinâmica são poucos, e a maioria é para interseções, rodovias e redes pequenas”.

A segunda característica distintiva está relacionada com a maneira na qual a informação prévia é avaliada. Alguns métodos expressam isso em forma de uma matriz O/D “alvo”, enquanto outros estimam o número de viagens atraídas ou produzidas por cada zona. Informações sobre a matriz O/D alvo podem ser obtidas por uma pesquisa amostral. Como Abrahamsson (1998, tradução nossa) coloca:

em abordagens estatísticas, a matriz O/D alvo é considerada pertencente a alguma distribuição estatística e pode ser obtida estimando os parâmetros de distribuição estatística.

Em abordagens baseadas em modelo de tráfego, a matriz O/D alvo é assumida normalmente por uma matriz O/D anterior e pode ser ajustada para satisfazer as contagens de tráfego.

Essa matriz prévia – às vezes chamada de matriz semente – fornece um ponto de partida para a procura de solução iterativa. Se o comportamento de viagem é desconhecido, a matriz prévia de demanda pode conter dados irrealistas (todas as células com a mesma demanda). O modelo modifica sucessivamente as magnitudes das células sementes inicial, multiplicando-as por um fator de correção. Em ambas as abordagens, os algoritmos são utilizados para minimizar a distância entre os valores da matriz observada e da estimada. Estas abordagens estão minuciosamente descritas na subseção 3.2.

2.1 Riscos e problemas de estimação da matriz a partir de contagens de tráfego

A estimação da matriz O/D seguindo uma abordagem empírica *bottom-up* implica alguns riscos. Primeiramente, modelos construídos nessa base não têm validade para a análise de longo prazo, uma vez que há perda do histórico comportamental. Segundo, geralmente se assumem os dados de contagem de tráfego como “verdade”, enquanto eles estão sempre sujeitos a incerteza associada à época da pesquisa e com o procedimento de amostragem e o desenho da pesquisa. Como Ortúzar e Willumsen (2011, tradução nossa) enfatizam: “Um aspecto particular que é preciso ter em mente é o fato de que as técnicas de estimativa de matriz podem tentar forçar uma matriz de viagens ajustada para reproduzir contagens de tráfego, mesmo se houver erros significativos na rede, no modelo de alocações ou nas próprias contagens.”

Procedimentos de estimação são dependentes da qualidade dos dados recolhidos pela contagem. Ortúzar e Willumsen (2011) referem-se a duas propriedades das contagens: independência e consistência. Ambas estão relacionadas com a continuidade do fluxo esperado em *links* contíguos da rede. No entanto, em situações reais, essa premissa é raramente atendida. Portanto, os autores recomendam o uso de “uma técnica para remover as inconsistências de fluxo de *links* nas contagens, a fim de assegurar que as condições de continuidade de fluxo entre *links* sejam atendidas” (Ortúzar e Willumsen, 2011, tradução nossa).

De acordo com Kuwahara e Sullivan (1987), os procedimentos de estimação de toda matriz O/D a partir dos dados de contagem deve lidar com uma série de questões:

- problema de *dupla contagem*: viagens de longa distância podem passar por mais de uma estação de pesquisa. A consequência é que certas viagens podem ser amostradas e ampliadas mais de uma vez;
- problema da *baixa cobertura geográfica*: escolhas de rotas menos usadas podem levar a subestimar certos pares O/D;

- utilização eficiente dos dados: fatores de expansão devem ser ajustados para compensar a dupla contagem e a baixa cobertura geográfica; e
- consequências da incerteza e comportamento de viagem desconhecido: a expansão da amostra baseia-se no uso de dados de entrada errados e suposições questionáveis. Efeitos de erros de entrada podem ser amplificados.

Podemos completar a lista anterior incluindo um quinto problema: *perda de informação*. A calibração da matriz por meio de contagens de tráfego se beneficia do fato de usar um dado empírico adicional, permitindo obter, no ano-base, fluxos mais confiáveis na rede. No entanto, em modelos de previsão, há também uma perda de informações causada pelo emprego de procedimentos de correção da matriz. Os dados adicionais sobre os fluxos de tráfego contados ajustam a matriz de demanda atual, mas não os fatores comportamentais subjacentes utilizados para determinar os valores da matriz em cenários alternativos ou em anos futuros. Como consequência, o modelo perde a sua sensibilidade para as mudanças de variáveis estruturais utilizadas no modelo de demanda.

2.2 Métodos

Abrahamsson (1998) e Bera e Rao (2011) apresentam, partindo de dados de contagem de tráfego, procedimentos diferentes de estimação da matriz O/D. Esta subseção enfoca as principais técnicas de estimação e as descreve brevemente. Para tanto, deixamos de lado os procedimentos de formulações de modelagem sofisticados, como o modelo *logit* aninhado, apresentado por Hsiao e Hansen (2011). Esses autores estimam um conjunto de parâmetros relacionados com muitas decisões no contexto de um modelo de demanda para transporte aéreo. O conjunto alternativo para rotas inclui um voo direto da origem ao destino, bem como as rotas que envolvem conexões em *hubs*. No entanto, os autores usam dados de cinco fontes, e não apenas dados de contagem. A seguir, as principais técnicas de estimação.

2.2.1 Métodos baseados em modelos de demanda de viagens: modelos gravitacional e gravitacional de oportunidade

A matriz O/D pode ser um resultado de um modelo gravitacional com base em contagens de tráfego e em suposições a respeito de potencialidades de geração e atração de cada zona. A ideia básica por trás desse tipo de modelo é construir e calibrar um modelo gravitacional, levando em conta fatores como o tamanho da população e o emprego, entre outros dados zonais. Nesse modelo, os volumes entre um par O/D são a variável dependente a ser encontrada. Embora Ortúzar e Willumsen (2011) declararem a escolha entre a abordagem gravitacional ou gravitacional de oportunidade ser algo a ser decidido empiricamente, eles relatam que o modelo gravitacional dá o melhor ajuste no nível de matriz de viagens.

2.2.2 Abordagem de maximização da entropia

O procedimento de maximização de entropia procura descobrir a matriz viagem mais provável com informação disponível (por exemplo, uma matriz prévia ou suposições sobre a distribuição do tempo – impedância – da viagem) por um método iterativo. A desvantagem desse procedimento é que ele não considera as incertezas nas contagens de tráfego. Esse método é equivalente à abordagem de minimização de informação.

2.2.3 Abordagens estatísticas

Técnicas de estimação paramétricas podem também ser empregadas. Máxima verossimilhança implica fazer suposição da distribuição de viagens no conjunto de contagens de tráfego observadas. Essa abordagem parte de dados de contagens de tráfego observados em um conjunto de amostra de fluxos O/D, no qual é assumido seguir uma distribuição conhecida. Inferências a partir dos quadrados mínimos generalizados contam com o teorema de Gauss-Markov e não requerem hipóteses sobre a distribuição estatística dos dados de contagem de tráfego. O método de inferência bayesiana considera a matriz O/D alvo como uma função de probabilidade prévia da matriz a ser estimada.

3 RECOMENDAÇÕES DE MODELAGEM

Nós não conseguimos encontrar, por meio da nossa pesquisa de literatura, relatórios detalhados sobre a aplicação prática de um modelo operacional de transporte aéreo de passageiros de longa distância, embora o tema tenha sido objeto de um programa europeu de pesquisa (PDC, 2000). Dessa forma, as recomendações incluídas neste relatório são derivadas das considerações teóricas antes resumidas e da estrutura dos dados empíricos que estarão disponíveis.

Partindo de questões discutidas na seção anterior, podemos afirmar o seguinte sobre os conjuntos de dados disponíveis.

- 1) Os dados de demanda primários são aqueles fornecidos pela Agência Nacional de Aviação Civil (Anac). A agência reguladora estatal registra em uma base mensal a quantidade de viagens aéreas por cada rota com partida ou chegada a um aeroporto brasileiro relevante, a partir dos registros de passageiros dos voos comerciais. Dados da Anac fornecem uma base empírica confiável e abrangente para estimar os fluxos de tráfego aéreo. Nesse caso, podemos supor que a contagem de dados é “verdadeira”, ou a melhor representação disponível dos deslocamentos reais no nível de *link* (par de aeroportos, o que é diferente dos pares O/D, visto que uma viagem pode fazer uso de conexões em aeroportos *hub*).
- 2) Além disso, uma pesquisa O/D de passageiros foi realizada pela EPL nos aeroportos. Em uma amostra de usuários, são feitas perguntas sobre a origem e o destino final da viagem; tempo, custos e modo de transporte para acessar e deixar os aeroportos; propósito de viagem; razões de rota e escolha do modo; bem como dados socioeconômicos.
- 3) Com base em dados georreferenciados em unidades espaciais (microrregiões do Instituto Brasileiro de Geografia e Estatística – IBGE), locais de aeroportos e rotas de avião, um modelo de rede para o modo aéreo contendo zonas de tráfego, aeroportos, linhas e rotas entre aeroportos pode ser construído com *software* de modelagem de transporte apropriado. Essa rede básica representará os serviços de transporte aéreo nacional disponíveis no ano-base e pode ser desenvolvida com dados de frequência e horários precisos.

3.1 Preparação dos dados/sistema de zoneamento

Por um lado, *para alguns pares O/D, pode não haver ligações aéreas diretas*. Possibilidades para lidar com este problema são: ignorar as células sem aeroportos, ou considerar uma ligação para – ou a partir dessas – essas células por um modo secundário até as zonas com aeroportos.

Por outro lado, as zonas densamente povoadas – por exemplo, grandes regiões metropolitanas, como São Paulo e Rio de Janeiro – podem conter mais de um aeroporto. Esses nós de transporte devem atrair os passageiros de quase a mesma área se os voos têm os mesmos destinos. Se não for possível ou desejável *definir uma rede na qual cada zona contém não mais do que um aeroporto, um modelo de escolha de aeroporto deve ser adicionalmente desenvolvido*. Uma possibilidade para tratar esses casos, em nível zonal, é adicionar na estrutura geral de modelagem um *modelo de escolha do aeroporto, a fim de representar a escolha de partidas e chegadas de passageiros dos aeroportos*. A atratividade do aeroporto deve refletir o custo direto, sua conveniência em termos de tempo de acesso (saída) de (para) os principais destinos e o nível de serviço, entre outros fatores (Lieshout, 2012). O desenvolvimento desse modelo pode apoiar-se no desenho de cenários para a análise de decisões de investimento para expansão da capacidade em nível intrazonal, uma vez que a situação de maior gargalo atualmente no transporte de passageiros brasileiro ocorre nas regiões metropolitanas.

3.2 Considerações metodológicas e opções

A estimativa do tráfego de passageiros pode ser estática (abordagem de alocação proporcional) ou dinâmica (alocação não proporcional, considerando os efeitos de congestionamento). A escolha da abordagem apropriada depende do objetivo do modelo e do tempo de resolução do modelo. Algumas ligações aéreas têm uma grande parcela da demanda de passageiros, e preços de tarifas refletem parcialmente a relação entre oferta e demanda. A escolha de um modelo dinâmico poderia ter em conta, por exemplo, que as viagens de Brasília para São Paulo tendem a ser mais caras na sexta-feira à tarde, como consequência da maior demanda de executivos e empresários. A fim de economizar dinheiro, algumas pessoas que querem viajar de Brasília para São Paulo podem considerar mudar o horário da partida do voo, pegando outras rotas – por exemplo, com conexões, aceitando maiores tempos de viagem –, ou mesmo escolhendo outros modos de transporte (ônibus ou carro). *Supondo-se que as proporções de viajantes que escolhem uma rota não dependem de volume de tráfego sobre ele* – isto é, os efeitos de congestionamento diários são insignificantes, uma vez que o projeto tem como objetivo simular efeitos de longo prazo e, portanto, lida com prazos mais amplos –, *o procedimento de modelagem estática parece ser mais apropriado*. A escolha do horário de partida e a interdependência entre o nível de demanda e a diferenciação de preços não devem ser tratadas pelo modelo concebido para estimar os fluxos em uma base anual ou mensal. Consequentemente, as proporções de escolha de rota são independentes do processo de estimação da matriz O/D.

Os pesquisadores vêm estudando os problemas de estimação em sua maioria voltados para o transporte rodoviário de passageiros. Vários esforços são direcionados para a inferência estatística da matriz a partir de uma amostra limitada de observações no nível da ligação e o problema de escolha de rota. Para o modo de carro, ainda é muito difícil reunir dados consistentes de todas as ligações, apesar de um tremendo progresso técnico. No entanto, *verificar as condições de continuidade de fluxo de ligação* não é um grande problema para o transporte aéreo de passageiros, uma vez que existe um número limitado de rotas que ligam os nós ou as zonas. Nem o problema de dupla contagem é grave, porque cada contagem representa, de fato, um passageiro que viajou no *link*, e todos os *links* (rotas de voos comerciais) são considerados nos registros da Anac.

3.3 Modelo de geração/distribuição

Um modelo de geração conjuntal/distribuição deve ser desenvolvido, a fim de garantir a produção de matrizes de origem-destino realistas – sem superestimar quantidades em zonas com aeroportos hubs importantes.

O número de partidas (chegadas) de voos em aeroportos com função de *hub* pode ser desproporcionalmente maior do que o potencial de geração (atração) das zonas de tráfego nesses nós. *As variáveis socioeconômicas no nível de zona podem ser incluídas em uma função para fornecer as quantidades que devem ser encontradas pela soma dos fluxos gerados em cada origem ou atraídos por cada destino.* Uma vez que nenhuma matriz O/D alvo está disponível, o modelo de demanda de transporte do tipo gravitacional, sugerido por Ortúzar e Willumsen (2011), com base nas comparações entre especificações alternativas tratadas por Tamin e Willumsen (1989), parece ser apropriado. Ao considerar os fatores de produção e atração relevantes, esse método deve assegurar que uma proporção considerável dos volumes de ligações que parte de uma zona populosa chega à outra zona populosa. De uma forma muito simples, o modelo gravitacional pode levar em conta o número de chegadas e partidas de avião em cada zona de tráfego. Estes totais podem ser calculados diretamente a partir dos dados fornecidos pela Anac. Em uma versão posterior, podem-se incluir, na fórmula de geração, variáveis que podem ser previstas no futuro.

Como exposto anteriormente, um problema relevante com os dados é a falta de informação real sobre as origens ou os destinos da cadeia de viagens. Por exemplo, alguém voando de Salvador para Belo Horizonte pode na verdade estar fazendo uma viagem de Fortaleza para São Paulo com duas escalas. As informações sobre o ponto de partida e o destino pretendido não estão disponíveis nos dados da Anac, e este passageiro pode ser contado três vezes – como passageiros de três voos – no nível do fluxo de ligação do aeroporto. Os dados da pesquisa O/D de passageiros aéreos da EPL podem ser usados para transformar a matriz de fluxo de aeroporto (Anac) em uma matriz O/D aproximada baseada em zonas de tráfego. *Dados de pesquisa O/D de passageiros podem ser usados para tratar volumes de aproximação e partidas das zonas O/D.* A quantidade de viagens em que os aeroportos de partida e chegada não coincidem com pontos inicial e final da cadeia de viagem deve ser deduzida dos totais de partidas e chegadas calculados pela simples adição de volumes nos *links* (registros da Anac). Em outras palavras, os dados de pesquisa podem ser utilizados num passo de correção de quantidades de produção e de atração no modelo de geração. Expandindo as informações de pesquisa e ajustando o total de viagens obtidas dos registros da Anac, essa é uma maneira de lidar com esse problema para todas as zonas, se os dados de pesquisa forem representativos e confiáveis.

3.4 Modelo de alocação

Para muitos pares O/D – em particular para as zonas vizinhas de baixa ou média densidade –, espera-se que apenas uma rota de ligação de avião esteja disponível. Como consequência, a “escolha de rota” é um problema simplificado ou é um caso particular da técnica de alocação proporcional, na qual a proporção de viagens entre as zonas i e j , utilizando o *link* a , muitas vezes é igual a 1. Segue-se que, do ponto de vista metodológico, é semelhante à técnica de alocação “tudo ou nada”, pela qual todo o tráfego é atribuído à rota com custo de viagem mais baixo (generalizada). *Se todos os viajantes só escolhessem um link direto para viajar de certo par O/D, o tráfego aéreo de passageiros ao longo de um determinado caminho poderia ser considerado igual à quantidade de viagens neste par O/D.*

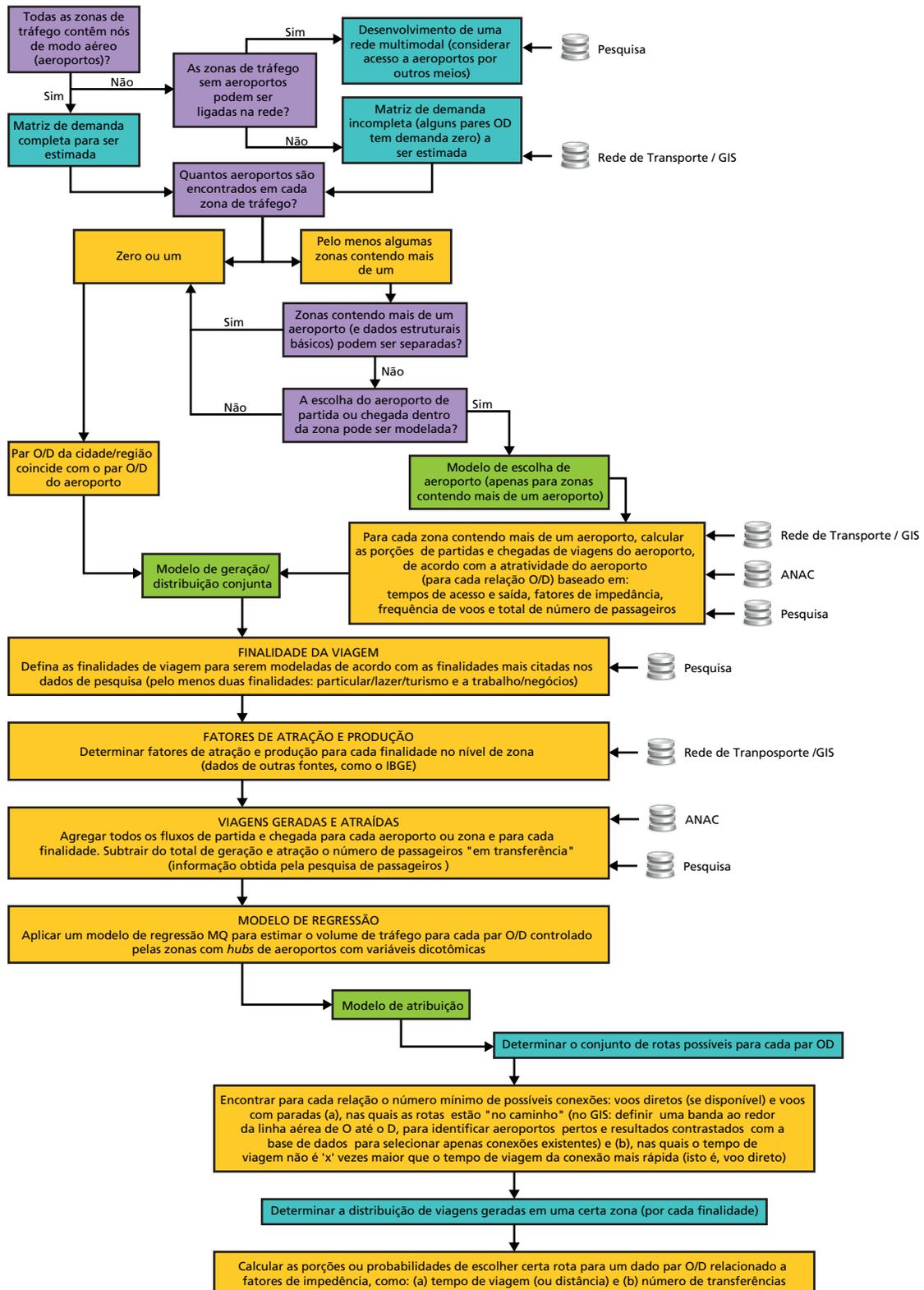
Apesar disso, é irrealista assumir que, para todas as relações de O/D dadas, os viajantes escolheriam apenas o voo direto. Podem haver, para muitos pares O/D, várias possibilidades de conexão fora o voo direto – por exemplo, voo com paradas em vários outros nós. Apesar de geralmente levar mais tempo, estas alternativas podem ser atraentes para os segmentos que procuram opções mais baratas ou para aqueles que desejam fazer paradas por outras razões. A possibilidade de escolher muitas rotas para viajar entre um par de zonas O/D gera um problema matemático, uma vez que o fluxo de ligação direta não expressa necessariamente o potencial de atração e produção das respectivas zonas. Portanto, o conjunto de caminhos viáveis para cada par O/D será determinado com base em dados empíricos. Para determinar isso, *algoritmos usando bases de dados georreferenciados podem ser implementados por meio da identificação de um conjunto de rotas alternativas com tempos de viagem comparáveis à rota direta*. Isso pode ser feito de duas maneiras: *i)* definindo um corte relativo para cada par O/D (tempo de viagem não é maior do que ‘x’ vezes o mínimo de tempo de viagem observada); ou *ii)* listando e ordenando as conexões de acordo com os tempos de viagem, pegando as “n” conexões mais atraentes em termos de tempo de viagem.

Para vários pares origem-destino, haverá diferentes caminhos. Portanto, o modelo pode conter um *modelo de alocação* em que a *escolha de rota ou probabilidades pode estar relacionada a fatores de impedância como o tempo de viagem – ou distâncias – e número de transferências necessárias para viajar nesse link*. Fatores de impedância são derivados do modelo de rede e não são dependentes do nível de congestionamento.

O volume observado em cada *link* pode ser subdividido em dois componentes: o volume de viagens diretas sem paradas (ligando um par origem-destino “real”) e o volume de viagens com uma ou mais paradas. Este último componente é em si uma soma de volumes de viagens de origens reais (chegadas a um destino intermediário), volumes de viagens para destinos reais (partidas de uma origem intermediária) e volumes de viagens entre duas paradas. Se estivermos interessados em estimar o volume total para um par O/D, precisamos somar o volume de viagens diretas ligando este par O/D particular com outro componente. Esse componente pode ser tanto o volume de viagens partindo da origem com paradas em outras zonas, quanto determinado destino ou o volume de viagens para esse destino.

Uma visão esquemática geral sobre as etapas de modelagem pode ser realizada por intermédio da figura 1.

FIGURA 1
Esquema de modelagem geral sugerida



Elaboração dos autores

4 INÍCIO DE APLICAÇÃO DA PROPOSTA: DESENVOLVIMENTO DO MODELO DE GERAÇÃO DE VIAGENS AÉREAS DE PASSAGEIROS

Esta seção relata o desenvolvimento do modelo de geração de viagens aéreas de passageiros, a partir dos dados de viagens observadas obtidos dos registros da Anac entre os pares de aeroportos brasileiros. Trata-se de um modelo econométrico, desenvolvido para estimar os volumes produzidos e atraídos por microrregiões, calibrado para o ano-base 2010 e com capacidade preditiva para os anos futuros da simulação.

4.1 Dados

Foi estruturado um banco de dados contendo informações de diversas fontes sobre as microrregiões brasileiras. Microrregiões são unidades territoriais oficialmente definidas pelo IBGE e que oferecem uma escala geográfica apropriada para a definição das *zonas de tráfego* do modelo de transporte aéreo de passageiros. Informações disponíveis para unidades territoriais menores – como o volume de viagens em aeroportos, inicialmente associado a municípios – foram agregadas no nível da microrregião.

O banco de dados contém um conjunto de variáveis estruturais – como população, renda, emprego etc. – obtidas junto a fontes estatais (essencialmente, IBGE e Ipea), para serem utilizadas como preditores dos modelos econométricos. O banco de dados também tem informações sobre a demanda por transporte aéreo (volume de passageiros) observada em cada aeroporto ou par de aeroportos. Os volumes de passageiros pagantes, originalmente disponibilizados pela Anac por mês, foram agregados em uma base anual, de modo que oscilações mensais ou sazonais da demanda possam ser desconsideradas. Adicionalmente, foram introduzidas informações produzidas a partir de análises conduzidas pela própria equipe do projeto de pesquisa.

De particular importância para o modelo de geração, foram geradas variáveis categóricas sobre a posição hierárquica de cada aeroporto em uma rede *hub-and-spoke*. Visto que não há uma hierarquia oficial estabelecida para os aeroportos brasileiros, a posição hierárquica de cada aeroporto foi definida a partir de uma análise com base nos dados de demanda e no número de relações com outros aeroportos. A cada microrregião foi associado o maior nível hierárquico de aeroportos contidos em seu território. As microrregiões São Paulo (SP), Rio de Janeiro (RJ), Guarulhos (SP) e Brasília (DF) estão relacionadas ao primeiro nível hierárquico. No segundo nível, encontram-se Campinas (SP), Curitiba (PR), Recife (PE), Fortaleza (CE), Salvador (BA), Porto Alegre (RS) e Belo Horizonte (MG). Estas são seguidas, no terceiro nível hierárquico, por outras onze microrregiões, conforme mostra a tabela 1.

TABELA 1
Posição hierárquica das microrregiões na rede *hub-and-spoke* de transporte aéreo de passageiros

Microrregião	População (1.000)	Viagens (1.000)	Hub	Microrregião	População (1.000)	Viagens (1.000)	Hub
Rio de Janeiro	10.118,79	10.733,86	1	Vitória	1.342,021	1.683,66	3
Guarulhos	1.141,555	10.074,22	1	Belém	1.805,616	1.557,196	3
São Paulo	12.002,96	8.326,938	1	Florianópolis	771,384	1.428,23	3
Brasília	2.180,376	8.031,834	1	Goiânia	1.812,128	1.390,35	3
Belo Horizonte	4.140,128	5.029,69	2	Manaus	1.654,578	1.379,925	3
Campinas	2.295,561	4.713,467	2	Cuiabá	721,306	1.293,373	3
Salvador	2.987,508	4.316,058	2	Natal	887,237	1.200,773	3

(Continua)

(Continuação)

Microrregião	População (1.000)	Viagens (1.000)	Hub	Microrregião	População (1.000)	Viagens (1.000)	Hub
Porto Alegre	3.149,478	3.590,008	2	São Luís	1.105,717	939,669	3
Curitiba	2.628,289	3.317,229	2	Arapiraca	336,385	808,8	3
Recife	2.805,512	3.062,271	2	Maceió	954,631	803,811	3
Fortaleza	2.858,127	2.765,547	2	Foz do Iguazu	347,048	786,76	3

Fontes: IBGE; Anac.
Elaboração dos autores.
Obs.: A tabela 2 descreve a fonte de dados para cada variável.

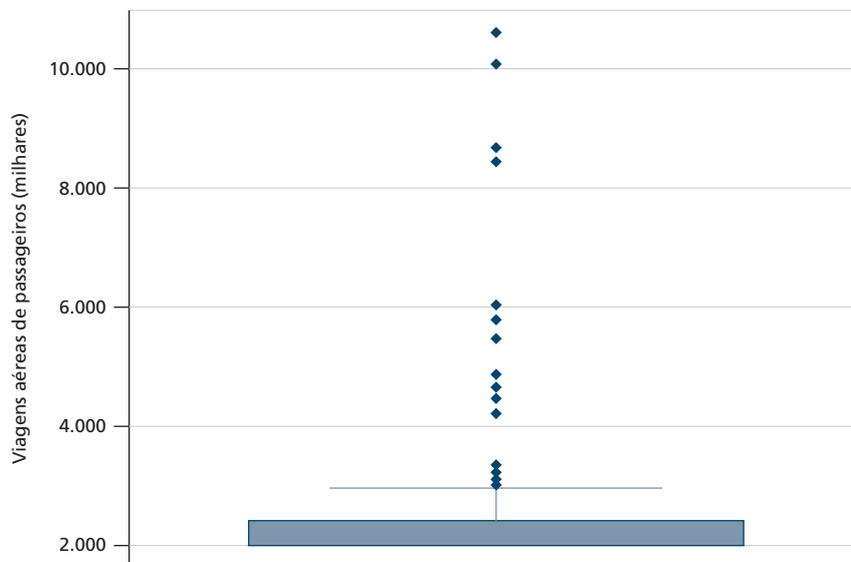
É importante frisar que grande parte das 554 microrregiões não pode ser considerada nos modelos de regressão, visto que, ao não conter aeroportos em seu território, lhes faltam informações sobre a demanda. A tabela 2 informa medidas descritivas e a fonte das principais variáveis consideradas para os modelos de demanda de 115 microrregiões, para as quais há informações completas disponíveis.

TABELA 2
Estatísticas descritivas básicas das variáveis do modelo de geração (2010 e 2012)

Variável	Significado	Ano	Fonte	Média	Desvio-padrão	Mediana	Mínimo	Máximo
volm	Número de passageiros embarcados (milhares)	2012	Anac	757,46	1.893,55	58,59	0,09	10.733,86
popm	População (milhares)	2010	IBGE	765,26	1.558,86	320,84	2,26	12.002,96
pi bpm	Rendimento <i>per capita</i> (milhares)	2010	IBGE	186,09	168,41	127,94	12,79	687,93
empm	Número de empresas (milhares)	2010	IBGE	28,81	71,97	10,98	0,23	671,94
vclm	Número de automóveis particulares (milhares)	2010	IBGE	239,84	641,10	71,02	0,10	5.902,67
sus	Número de unidades de saúde básica (milhares)	2010	IBGE	198,50	183,18	148,00	3,00	1.087,00

O gráfico tipo *box-plot* da variável dependente do modelo (o volume de viagens aéreas de passageiros) indica uma considerável quantidade de observações com valores acima do percentil 75. Para a modelagem econométrica, o número relativamente alto de microrregiões com elevados volumes gerados dá suporte à introdução de variáveis *dummy* sobre os mais altos níveis hierárquicos dos aeroportos na rede *hub-and-spoke*.

GRÁFICO 1
Box-plot do número de passageiros embarcados por microrregião



Fonte ANAC (2012).
Elaboração dos autores.

4.2 Modelagem econométrica

Com base nas variáveis disponíveis, foram desenvolvidas duas estratégias metodológicas. A primeira consiste em adotar, sequencialmente, um modelo de geração e um modelo de distribuição de viagens. A segunda trata-se de aplicar um modelo de demanda direta. Ao final de ambos os processos, obtém-se uma matriz origem-destino para o transporte aéreo de passageiros no Brasil no ano-base de 2010, que pode ser utilizada para a predição dos anos futuros da simulação.

Os modelos de demanda direta visam estimar simultaneamente mais de uma das quatro etapas do método tradicional. Neste caso, o modelo combina as etapas de geração e distribuição, tendo como variável dependente o volume de passageiros de uma zona de origem para uma zona de destino. Conforme a literatura, esses *modelos de interação* estimados por mínimos quadrados ordinários (MQO) são considerados um tipo de modelo econométrico espacial. Matematicamente, o modelo de demanda direta acopla uma função de impedância aos fatores responsáveis pela geração e atração de viagens utilizados pela metodologia tradicional.

Na série de modelos desenvolvidos, foram testadas diversas funções de impedância, todas elas baseadas na distância euclidiana entre os centroides das microrregiões calculada com o *software* de georreferenciamento QGIS. Outras medidas de impedância, como o custo monetário do transporte aeroviário – a exemplo do custo tarifário ou do yield médio – para cada par origem-destino, não foram incluídas devido à falta de acesso aos dados. Além disso, as especificações dos modelos de demanda direta variaram com relação às variáveis explicativas introduzidas. No entanto, os resultados desses modelos foram considerados insatisfatórios em função da baixa capacidade explanatória e preditiva (em geral, R^2 ajustado abaixo de 0,40).

Os modelos tradicionais de geração de viagens visam estimar o número total de viagens produzidas e atraídas por cada zona de tráfego – ou seja, a variável dependente é diferente da do modelo de demanda direta. Resultantes desses modelos são vetores com a quantidade de viagens produzidas ou atraídas, posteriormente relacionados entre si por meio de um algoritmo de distribuição – não detalhado neste relatório –, produzindo as matrizes origem-destino. De acordo com a literatura, o tamanho populacional, a renda e o tamanho da frota de veículos são alguns dos fatores tradicionalmente considerados para elaborar estimativas dos volumes de tráfego gerados em cada região.

4.2.1 Análise de correlação e multicolinearidade

Antes de apresentarmos os resultados do modelo, é importante assinalar que, dada a natureza dos dados – sobre os quais não se tem controle –, algumas hipóteses do modelo clássico de regressão linear podem ser eventualmente violadas. De particular importância no contexto de desenvolvimento dos modelos de geração baseados em variáveis estruturais das zonas de tráfego é a questão da multicolinearidade, uma vez que diversas variáveis incluídas são altamente correlacionadas entre si. Os coeficientes de correlação superam certas vezes o valor de 0,80 (tabela 3), como ocorre com as variáveis população, número de empresas e frota de veículos.

TABELA 3
Correlação entre as variáveis do modelo

	volm	popm	piGPCM	emPM	vclm	susm
volm	1,0000	–	–	–	–	–
popm	0,8033	1,0000	–	–	–	–
piGPCM	0,1434	0,2006	1,0000	–	–	–
emPM	0,7156	0,9456	0,2447	1,0000	–	–
vclm	0,7357	0,9479	0,2390	0,9965	1,0000	–
susm	0,6719	0,8413	0,3641	0,7412	0,7400	1,0000

Elaboração dos autores.

Obs.: Variáveis: volm = volume de passageiros aéreos originados na zona; popm = população da zona; piGPCM = produto interno bruto (PIB) *per capita* da zona; emPM = número de empresas da zona; vclm = frota de veículos da zona; sus = número de estabelecimentos do Sistema Único de Saúde (SUS) da zona.

Para a estimação dos coeficientes em um modelo de regressão, assume-se que os regressores sejam independentes entre si e haja suficiente variabilidade em seus valores. Quando os preditores são altamente colineares, é difícil isolar o impacto individual de cada regressor sobre o regressando. Os coeficientes de regressão podem ser determináveis – se não existir correlação perfeita entre os preditores –, mas apresentarão elevados erros-padrão, não podendo ser estimados com exatidão.

Sob multicolinearidade, ainda que os coeficientes estimados pelo método de mínimos quadrados ordinários possam ser considerados os melhores estimadores não viesados e eficientes, eles apresentam maiores variâncias e covariâncias, implicando intervalos de confiança mais amplos e razões *t* insignificativas, indicando que os coeficientes não são estatisticamente diferentes de zero. Embora a razão *t* de um ou mais coeficientes seja estatisticamente insignificante, o valor da medida global de ajuste do modelo (R^2) pode ser bastante alto (Gujarati, 2000, p. 326).

Entretanto, nos modelos testados, os típicos sintomas de multicolinearidade discutidos na literatura não foram observados simultaneamente. Embora os valores dos coeficientes de determinação ajustados estejam acima de 0,95, sugerindo um excelente grau de ajuste geral, os estimadores da maior parte dos coeficientes permanecem estatisticamente significantes no nível de 1%.

Não há recomendações consensuais para a detecção ou para o tratamento da multicolinearidade – a literatura em econometria sugere a aplicação de uma série de métodos. Alguns autores sugerem que, mesmo quando detectada, o melhor tratamento para a multicolinearidade é não fazer nada, pois trata-se de um problema inerente aos dados com os quais se deve trabalhar. Assim, evitou-se descartar variáveis explicativas, o que poderia induzir a uma série de problemas mais sérios (estimadores viesados e inconsistentes dos coeficientes, variância dos erros incorreta, vies na estimação dos erros-padrão dos coeficientes etc.), causados pela omissão de variáveis relevantes.

Também foi testado um modelo de regressão baseado nos fatores extraídos de uma análise de componentes principais (*principal component analysis*). No entanto, esse procedimento, além de não apresentar resultados satisfatórios, inviabiliza uma leitura direta dos efeitos de determinadas variáveis sobre os volumes estimados.

4.2.2 Modelos estimados

Esta subseção apresenta a especificação e os resultados dos modelos clássicos de geração testados. Esses modelos foram estimados por MQO, utilizando-se o *software* Stata IC 13.1.

Como todos os modelos têm o volume produzido de viagens aéreas de passageiros como variável dependente e têm por base a mesma amostra, eles podem ser comparados com base no coeficiente de determinação (R^2 ou R^2 ajustado). Também é reportado para cada modelo o valor do critério de informação de Akaike, que penaliza modelos com muitas variáveis. A tabela 4 traz, para cada variável explicativa, o valor estimado do coeficiente, bem como o nível exato de significância (a estatística p , reportada entre parênteses).

Cumprir registrar que foram testadas diversas especificações alternativas do modelo clássico de geração – por exemplo, com transformação logarítmica das variáveis e introduzindo variáveis de interação. Estes não apresentaram vantagens evidentes e, por questões de espaço, não serão aqui discutidos. No entanto, para possibilitar uma comparação, os resultados de alguns desses modelos são reportados no apêndice.

TABELA 4
Modelos clássicos de geração estimados

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Modelo 7	Modelo 8	Modelo 9	Modelo 10
hub1	8.781 (0,000)	8.581 (0,000)	8.461 (0,000)	8.500 (0,000)	8.697 (0,000)	8.860 (0,000)	8.492 (0,000)	8.429 (0,000)	8.749 (0,000)	8.974 (0,000)
hub2	3.588 (0,000)	2.827 (0,000)	2.452 (0,000)	2.936 (0,000)	3.481 (0,000)	3.380 (0,000)	2.721 (0,000)	2.877 (0,000)	2.936 (0,000)	3.411 (0,000)
hub3	1.126 (0,000)	792 (0,000)	751 (0,000)	889 (0,000)	1.020 (0,000)	1.009 (0,000)	879 (0,000)	871 (0,000)	882 (0,000)	1.023 (0,000)
popm	0,080 (0,018)	0,436 (0,000)	0,569 (0,000)	0,189 (0,138)	0,369 (0,000)	0,340 (0,000)	0,316 (0,024)	0,194 (0,003)	– (–)	0,305 (0,000)
popm2	– (–)	0,000 (0,000)	0,000 (0,000)	0,000 (0,899)	– (–)	– (–)	0,000 (0,531)	– (–)	– (–)	– (–)
pi bpm	– (–)	– (–)	-1,217 (0,005)	– (–)	– (–)	-0,492 (0,000)	-0,799 (0,049)	– (–)	– (–)	0,461 (0,001)
pi bpm2	– (–)	– (–)	0,002 (0,001)	– (–)	– (–)	– (–)	0,002 (0,008)	– (–)	– (–)	– (–)
empm	– (–)	– (–)	– (–)	4,309 (-0,131)	-6,412 (0,000)	-6,602 (0,000)	2,274 (0,449)	– (–)	– (–)	– (–)
empm2	– (–)	– (–)	– (–)	-0,012 (-0,001)	– (–)	– (–)	-0,010 (0,007)	– (–)	– (–)	– (–)
vclm	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	0,558 (0,035)	0,701 (0,005)	-0,680 (0,000)
vclm2	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	0,000 (0,000)	0,000 (0,000)	– (–)
sus	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	0,596 (0,008)	– (–)
R2	0,972	0,977	0,979	0,981	0,977	0,980	0,983	0,981	0,981	0,979
R2 ajustado	0,971	0,976	0,780	0,980	0,976	0,978	0,981	0,980	0,980	0,977
AIC	1.677	1.655	1.648	1.635	1.655	1.644	1.630	1.633	1.635	1.649

Elaboração dos autores.

Obs.: Variáveis: volm = volume de passageiros aéreos originados na zona; hub1 = variável *dummy* que indica que a zona possui um aeroporto *hub* de nível 1; hub2 = variável *dummy* que indica que a zona possui um aeroporto *hub* de nível 2; hub3 = variável *dummy* que indica que a zona possui um aeroporto *hub* de nível 3; popm = população da zona; popm2 = população da zona elevada ao quadrado; pi bpm = PIB *per capita* da zona; pi bpm2 = PIB *per capita* da zona ao quadrado; empm = número de empresas da zona; empm2 = número de empresas da zona ao quadrado; vclm = frota de veículos da zona; vclm2 = frota de veículos da zona ao quadrado; sus = número de estabelecimentos do SUS da zona.

Variável dependente = volm.

Com respeito às variáveis, merecem particular atenção duas características em comum dos modelos antes citados. Primeiramente, estes modelos foram estimados sem intercepto, refletindo a tese de que zonas inabitadas ou sem atividade econômica não devem produzir viagens. Essa é uma recomendação da literatura sobre modelagem de demanda de transportes. Como se espera que a linha da regressão estimada passe exatamente pela origem, modelos com interceptos significativamente distintos de zero devem ser rejeitados, enquanto modelos com interceptos iguais ou próximos de zero devem ser reestimados sem o termo de intercepto (Ortúzar e Willumsen, 2011, p. 152).

Uma segunda característica relevante é que os modelos apresentados contêm como preditores três variáveis *dummy*, indicando o nível hierárquico mais elevado de um aeroporto no território de uma dada microrregião. Compõe a categoria de referência o grupo de microrregiões sem aeroportos com a função de *hub*. Em todos os modelos, as três variáveis são significativamente diferentes de zero em um nível de confiança de 99%; conforme as expectativas, todas elas têm efeitos parciais positivos sobre o volume de viagens geradas. Quanto mais importante o *hub* na hierarquia aeroportuária nacional, maior o volume de passageiros transportados.

Além das variáveis categóricas sobre o nível hierárquico, a variável população é incluída em todos os modelos, com exceção do modelo 9. Os modelos 2, 3, 4 e 7 incluem como variável explicativa a população ao quadrado, mas só nos modelos 2 e 3 essa variável é estatisticamente significativa. Os modelos 3, 4, 6 e 10 incorporam também a variável produto interno bruto (PIB) *per capita*, estatisticamente significativa a 5% nessas especificações. Os modelos 4, 5, 6 e 7 trazem ainda a variável de emprego, que não apresenta o sinal esperado em duas especificações (modelos 5 e 6) e não é significativa a 10% em outras duas especificações (modelos 4 e 7). Os modelos 8, 9 e 10 têm a frota de veículos automotores como regressor. No entanto, o modelo 10 não expressa a esperada relação entre quantidade de veículos e número produzido de viagens, como se pode constatar pelo sinal negativo do coeficiente estimado. O modelo 9 traz, ainda, a variável sobre a quantidade de equipamentos de saúde básica. Os sinais negativos nos estimadores de emprego e número de veículos não necessariamente indicam inadequação do modelo. No caso da variável empregos, é possível que isso tenha ocorrido pela influência de zonas com relativamente elevada quantidade de pessoas aposentadas, mas com renda alta, o que elevaria o número de viagens aéreas não relacionadas ao número de empregos da zona. De forma semelhante, zonas com muitos veículos podem facilitar viagens de automóvel em detrimento das aéreas. Tal conclusão é corroborada pelo fato de o modelo 10 também incluir a variável *pibpcm*, esta sim com estimador positivo, o que indica uma possível compensação entre as duas variáveis.

Entre os modelos apresentados, o modelo 7 é o que apresenta os maiores R^2 e o mais baixo AIC. No entanto, os estimadores de duas variáveis não são significativos a 10%. Por isso, não consideramos que esse modelo seja suficientemente robusto para simular a demanda em anos futuros. O modelo 4 também pode ser descartado pela mesma razão. Outros modelos que apresentam excelentes graus de ajuste são os modelos 8, 9 e 10, embora o modelo 10 apresente um resultado contraintuitivo, talvez em função da alta correlação entre os regressores. Esse fenômeno – embora não se trate necessariamente de multicolinearidade – também pode estar presente nos modelos 8 e 9.

Expressamos preferência pelo modelo 3, que apresenta as seguintes características:

- os coeficientes de regressão estimados são estatisticamente significativos;
- o modelo apresenta um elevado grau de ajuste;

- o modelo é desenvolvido com um pequeno número de variáveis, relativamente fáceis de serem obtidas e projetadas em anos futuros; e
- as variáveis incluídas no modelo não são colineares (apresentam baixa correlação).

A tabela 5 apresenta de forma mais detalhada os resultados desse modelo.

TABELA 5
Resultados detalhados do modelo de regressão preferido

Fonte	SS	df	MS	Núm. observações =	115
Modelo	468.316.067	7	66.902.295	F(7,108) =	722,39
Resíduos	10.002.179	108	92.613	Prob > F	0,0000
Total	479.318.246	115	4.159.289	R2 =	0,9791
				R2 ajustado =	0,9777
				Raiz MSE =	304,32

volm	Coef.	Erro-padrão	t	P > t	(95% intervalo conf.)	
hub1	8.461	261	32,47	0,000	7.944	8.799
hub2	2.452	250	9,82	0,000	1.957	2.947
hub3	751	126	5,99	0,000	502	999
popm	0,569	0,108	5,260	0,000	0,354	0,786
popm2	0,000	0,000	-4,990	0,000	0,000	0,000
pibpscm	-1,217	0,425	-2,860	0,005	-2,059	-0,374
pibpscm2	0,002	0,001	3,290	0,001	0,001	0,004

Elaboração dos autores.

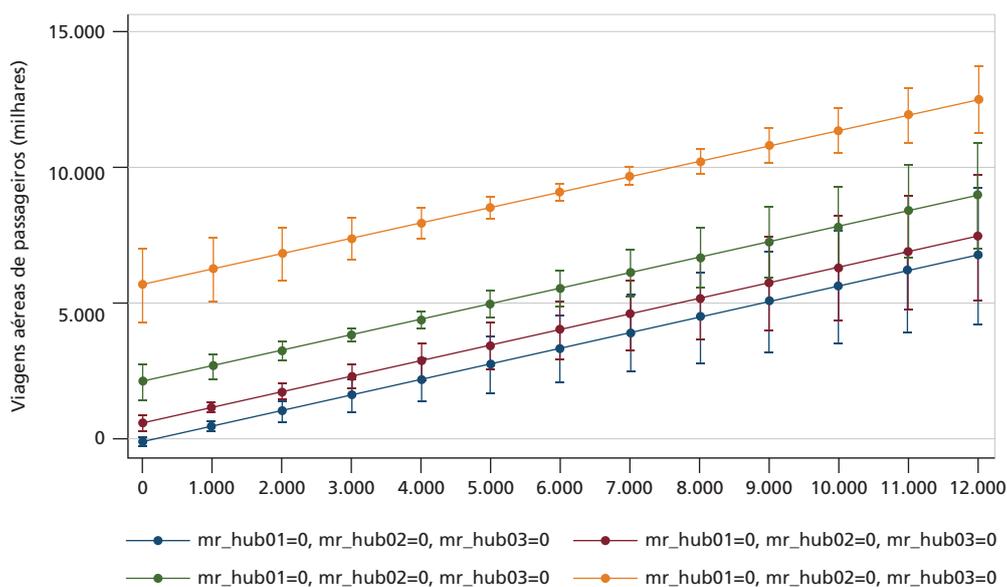
4.2.3 Interpretação dos resultados

O coeficiente de determinação (R^2) apresenta um valor de aproximadamente 0,98. Isso significa que cerca de 98% da variação do volume de viagens aéreas de passageiros é explicada pela variação das variáveis explicativas adotadas (*hub*, população e renda *per capita*). O modelo pode ser rodado sem termo de intercepto, forçando-o a passar pela origem.¹

Os coeficientes referentes aos *hubs* confirmam que microrregiões com aeroportos *hub* produzem um número maior de viagens aéreas em relação a microrregiões sem aeroportos *hub*, controladas as demais variáveis incluídas no modelo (população e renda *per capita*). As microrregiões que contêm um *hub* de terceira ordem – como Cuiabá, Manaus e Vitória – produzem, em média, 750,5 mil viagens a mais por ano. As microrregiões com *hubs* de segunda ordem contabilizam 2,45 milhões de viagens produzidas adicionais. Por fim, as microrregiões com os quatro mais importantes *hubs* aeroportuários nacionais têm 8,46 milhões de viagens a mais em relação a microrregiões sem *hubs*, tudo o mais constante. O gráfico 2 ilustra o volume adicional de viagens das microrregiões com as funções de *hub* de primeira (amarelo), segunda (verde) e terceira (vermelho) ordem para microrregiões de diferentes tamanhos populacionais, conforme os resultados da regressão rodada.

1. Pode-se constatar que a inclusão de um termo de intercepto no modelo com as mesmas variáveis independentes não aumenta seu poder explicativo. O intercepto é estatisticamente insignificante em um intervalo de confiança ao redor de zero.

GRÁFICO 2
Efeito de *hubs* sobre o volume de viagens com intervalo de confiança de 95%

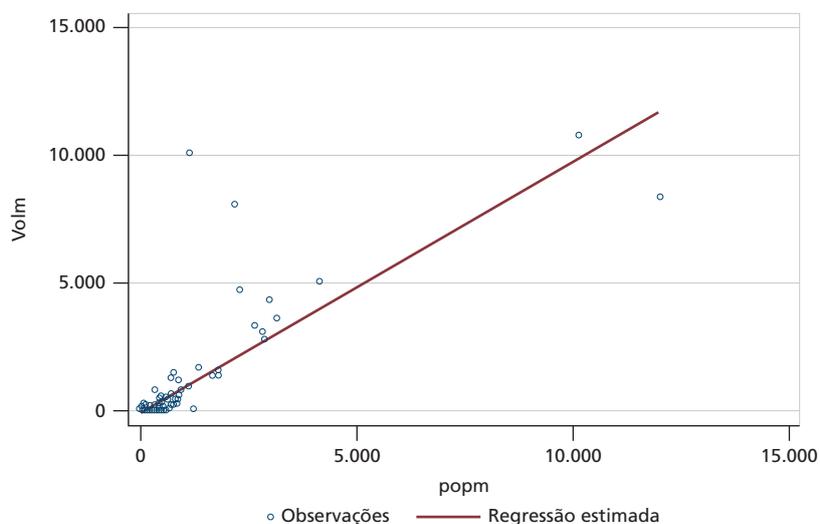


Elaboração dos autores.

As variáveis população e renda *per capita* mantêm uma relação polinomial de segundo grau com o volume de viagens produzido. O sinal do termo de primeiro grau é positivo, enquanto o de segundo grau é negativo (o valor do coeficiente $popm2$ é $-0,0000424$). Isso sugere que o volume de passageiros mantém uma relação crescente com o tamanho populacional, embora quanto mais populosa a microrregião, ligeiramente menor tende a ser sua contribuição marginal para o volume produzido de viagens.

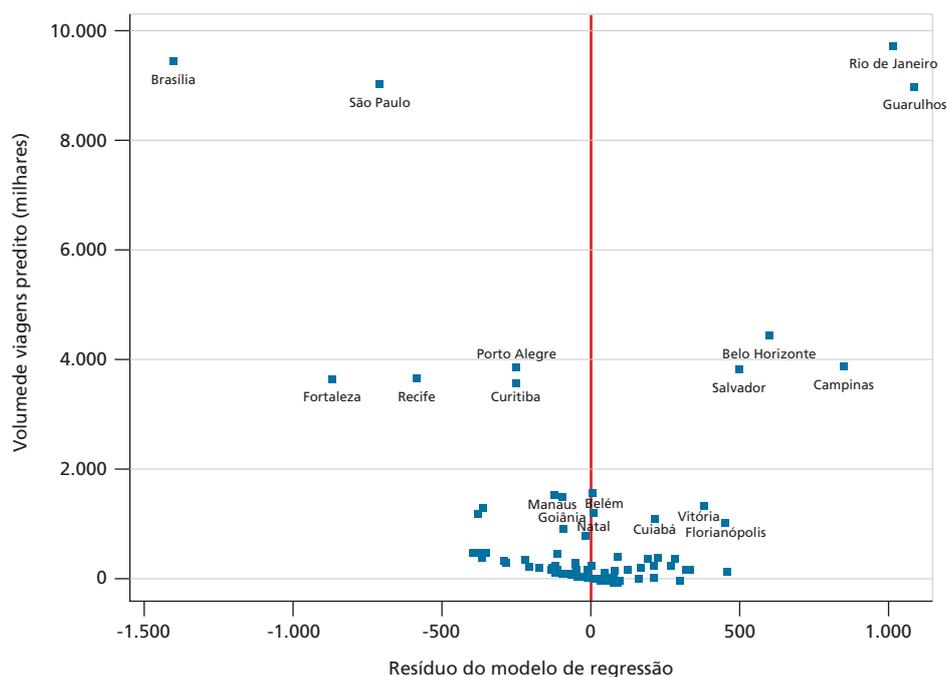
Analogamente, a variável PIB *per capita* mantém uma relação não linear com o volume de viagens. No entanto, seu sinal não condiz com a expectativa derivada da teoria. O termo de segundo grau é positivo, indicando que a intensidade da contribuição marginal do PIB *per capita* ao volume de viagens decai para regiões economicamente mais fortes.

GRÁFICO 3
Diagrama de dispersão do volume de viagens e população



Elaboração dos autores.

GRÁFICO 4
Resíduos e valores preditos



Elaboração dos autores.

Entretanto, o modelo é passível de críticas e eventualmente pode ser melhorado. Em primeiro lugar, como já ressaltado, maiores níveis de renda *per capita* estão associados a menores volumes de viagens aéreas. Embora não seja característica distintiva desse modelo no universo dos modelos testados, trata-se efetivamente de um resultado contraintuitivo. Em segundo lugar, embora o modelo produza estimativas próximas das efetivamente observadas para muitas microrregiões brasileiras, as predições tendem a ser menos acuradas em grandes centros urbanos. No gráfico 3, as observações posicionadas acima da linha vermelha (resíduo zero) estão subestimadas (o volume observado é maior que o volume estimado), enquanto observações abaixo da linha vermelha estão superestimadas. Os quatro mais importantes *hubs* do país apresentam relativamente grandes discrepâncias entre os volumes estimados e os concretamente observados. O gráfico 4 explicita o fato de que o modelo consegue prever melhor os volumes de microrregiões menos populosas e que geram menores volumes de viagens por avião. Tais resultados reforçam a necessidade de tratamento explícito da estrutura da rede de transporte aéreo no formato *hub-and-spoke*, em que os volumes de passageiros nos aeroportos *hub* são pouco influenciados pelas características socioeconômicas da zona em que estão localizados, mas sim pela quantidade e intensidade de ligações na malha aérea. Uma alternativa para melhorar a predictibilidade do modelo de regressão convencional de geração de viagens seria agregar variáveis socioeconômicas das regiões de influência de cada aeroporto. Entretanto, a não ser pelas zonas muito próximas dos aeroportos, a região de influência de cada aeroporto *hub* dependerá do desenho da malha aérea – ao menos, para os grandes *hubs* –, o que pode tornar o processo de predição igualmente difícil.

5 CONSIDERAÇÕES FINAIS

Este trabalho apresentou uma breve revisão bibliográfica dos modelos de estimação de matrizes O/D com base em contagens de tráfego. Em virtude de haver disponibilidade de dados de viagem em todos os *links* da rede de transporte aéreo brasileiro, propõe-se esse tipo de modelagem como adequado para a calibração de matrizes O/D de viagens aéreas de

passageiros. Entretanto, tal estimação não é direta, devido à estrutura da malha aérea no padrão *hub-and-spoke*. Para tanto, foi proposta uma abordagem para calibrar uma matriz O/D de transporte aéreo de passageiros, a partir dos dados que estarão disponíveis para a EPL; em especial, os registros de quantidades de passageiros mensais entre cada par de aeroportos da Anac e pesquisas O/D em aeroportos, que permitem identificar a origem e o destino reais de cada entrevistado, não apenas os aeroportos por onde passaram.

Como exercício, foi testada uma das etapas da abordagem proposta com base nos dados da Anac, a de modelagem de geração de viagens. Foi possível observar que a existência de aeroportos *hub* dificulta sobremaneira a capacidade preditiva de modelos de regressão convencionais. Isso reforça a utilização de modelos de estimação que absorvam as características da rede de transportes, como a estrutura *hub-and-spoke*. Espera-se, com a abordagem aqui proposta, contribuir para os trabalhos futuros de previsão de demanda por transporte aéreo, a fim de subsidiar as decisões públicas sobre os investimentos em infraestrutura aeroportuária e aérea.

REFERÊNCIAS

- ABRAHAMSSON, T. **Estimation of origin-destination matrices using traffic counts: a literature survey**. Laxenburg: IIASA, May 1998. p. 76. (IIASA Interim Report IR-98-021).
- BERA, S.; RAO, K. V. Estimation of origin-destination matrix from traffic counts: the state of the art. **European Transport/Trasporti Europei**, n. 49, p. 3-23, 2011. Disponível em: <<https://goo.gl/0seux8>>. Acesso em: 29 jun. 2014.
- FHWA – FEDERAL HIGHWAY ADMINISTRATION. **Improving existing travel models and forecasting processes: a white paper**. Alington: Federal Highway Administration, Dec. 2013. Disponível em: <<https://goo.gl/ZoFXnk>>.
- GUJARATI, D. N. **Econometria básica**. São Paulo: Makron Books, 2000.
- HSIAO, C.-Y.; HANSEN, M. A passenger demand model for air transportation in a hub-and-spoke network. **Transportation Research Part E: logistics and transportation review**, v. 47, n. 6, p. 1112-1125, Nov. 2011.
- KUWAHARA, M.; SULLIVAN, E. C. Estimating origin-destination matrices from roadside survey data. **Transportation Research Part B: methodological**, v. 21, n. 3, p. 233-248, June 1987.
- LIESHOUT, R. Measuring the size of an airport's catchment area. **Journal of Transport Geography**, v. 25, p. 27-34, Nov. 2012.
- ORTÚZAR, J. D.; WILLUMSEN, L. G. **Modelling transport**. 4. ed. Hoboken: John Wiley & Sons, 2011. p. 446.
- PDC – PETER DAVIDSON CONSULTANCY. **MYSTIC towards origin-destination matrices for Europe**. London: PDC, 2000.
- TAMIN, O. Z.; WILLUMSEN, L. G. Transport demand model estimation from traffic counts. **Transportation**, v. 16, n. 1, p. 3-26, Mar. 1989.

APÊNDICE

TABELA A.1
Outros modelos de regressão testados

	Modelo 1L	Modelo 2L	Modelo 3L	Modelo 4L	Modelo 5L	Modelo 6L	Modelo 7L	Modelo 8L	Modelo 9L	Modelo 10L	Modelo 11L	Modelo 12L	Modelo 13L
hub1	3,770 (0,000)	2,298 (0,037)	0,502 (0,670)	3,509 (0,001)	3,502 (0,001)	3,250 (0,002)	0,906 (0,429)	-0,048 (0,966)	0,502 (0,670)	1,872 (0,113)	2,416 (0,045)	2,242 (0,031)	-0,162 (0,898)
hub2	3,151 (0,000)	2,402 (0,003)	1,858 (0,024)	2,912 (0,001)	3,359 (0,000)	3,150 (0,000)	2,194 (0,007)	1,566 (0,052)	1,858 (0,024)	2,168 (0,007)	2,552 (0,002)	2,318 (0,003)	1,544 (0,067)
hub3	2,914 (0,000)	1,989 (0,004)	1,324 (0,057)	2,846 (0,000)	2,878 (0,000)	2,864 (0,000)	1,354 (0,054)	1,143 (0,093)	1,324 (0,057)	2,381 (0,000)	2,525 (0,000)	2,489 (0,000)	1,111 (0,109)
pop(log)	0,805 (0,000)	1,333 (0,000)	1,496 (0,000)	0,636 (0,008)	0,941 (0,000)	0,782 (0,002)	1,733 (0,000)	1,874 (0,000)	1,496 (0,000)	0,982 (0,000)	0,772 (0,000)	1,011 (0,000)	1,919 (0,000)
plbpc (log)	-	-0,560 (0,007)	-0,864 (0,000)	-	-	-	-0,854 (0,000)	-0,808 (0,000)	-0,864 (0,000)	-	-	-	-0,793 (0,000)
emp (log)	-	-	0,958 (0,002)	0,372 (0,598)	-	0,603 (0,394)	-	-	0,958 (0,002)	-	-	-	-0,190 (0,801)
sus (log)	-	-	-1,444 (0,001)	-	-0,834 (0,062)	-0,897 (0,048)	-1,363 (0,002)	-1,035 (0,019)	-1,444 (0,001)	-	-	-	-0,969 (0,045)
vd (log)	-	-	-	-0,114 (0,767)	0,222 (0,169)	-0,071 (0,851)	0,481 (0,004)	0,318 (0,061)	-	-	-	-	0,374 (0,307)
r_1sm	-	-	-	-	-	-	-	-9,778 (0,004)	-	-10,559 (0,004)	-	-11,519 (0,001)	-9,918 (0,013)
r_20sm	-	-	-	-	-	-	-	-	-	58,074 (0,506)	158,038 (0,058)	-	31,112 (0,742)
R2	0,973	0,974	0,977	0,973	0,974	0,974	0,977	0,979	0,977	0,975	0,974	0,975	0,979
R2 ajustado	0,972	0,973	0,976	0,971	0,972	0,972	0,975	0,977	0,976	0,974	0,972	0,974	0,977
AIC	476	471	461	480	476	478	463	456	461	468	475	466	490

Elaboração dos autores.

Obs.: Variáveis: vol (log) = logaritmo natural do volume de passageiros aéreos originados na zona; hub1 = variável *dummy* que indica que a zona possui um aeroporto *hub* de nível 1; hub2 = variável *dummy* que indica que a zona possui um aeroporto *hub* de nível 2; hub3 = variável *dummy* que indica que a zona possui um aeroporto *hub* de nível 3; pop (log) = logaritmo natural da população da zona; plbpc (log) = logaritmo natural do PIB *per capita* da zona; emp (log) = logaritmo natural do número de empresas da zona; vd (log) = logaritmo natural da frota de veículos da zona; sus (log) = logaritmo natural do número de estabelecimentos do SUS da zona; r_1sm = parcela da população com renda *per capita* abaixo de 1 salário mínimo (SM); r20sm = parcela da população com renda *per capita* superior a 20 SM.

Ipea – Instituto de Pesquisa Econômica Aplicada

EDITORIAL

Coordenação

Cláudio Passos de Oliveira

Supervisão

Everson da Silva Moura
Leonardo Moreira Vallejo

Revisão

Clícia Silveira Rodrigues
Idalina Barbara de Castro
Marcelo Araujo de Sales Aguiar
Marco Aurélio Dias Pires
Olavo Mesquita de Carvalho
Regina Marta de Aguiar
Reginaldo da Silva Domingos
Alessandra Farias da Silva (estagiária)
Lilian de Lima Gonçalves (estagiária)
Luiz Gustavo Campos de Araújo Souza (estagiário)
Paulo Ubiratan Araujo Sobrinho (estagiário)
Pedro Henrique Ximendes Aragão (estagiário)

Editoração

Bernar José Vieira
Cristiano Ferreira de Araújo
Daniella Silva Nogueira
Danilo Leite de Macedo Tavares
Jeovah Herculano Szervinsk Junior
Leonardo Hideki Higa
Herllyson da Silva Souza (estagiário)

Capa

Danielle de Oliveira Ayres
Flaviane Dias de Sant'ana

*The manuscripts in languages other than Portuguese
published herein have not been proofread.*

Livraria Ipea

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo
70076-900 – Brasília – DF
Tel.: (61) 2026 5336
Correio eletrônico: livraria@ipea.gov.br

Missão do Ipea

Aprimorar as políticas públicas essenciais ao desenvolvimento brasileiro por meio da produção e disseminação de conhecimentos e da assessoria ao Estado nas suas decisões estratégicas.

ipea Instituto de Pesquisa
Econômica Aplicada

MINISTÉRIO DO
**PLANEJAMENTO,
DESENVOLVIMENTO E GESTÃO**

