

# Produção de dados de tráfego sintéticos através de algoritmo genético e simulação microscópica

José Elievam Bessa Júnior<sup>1</sup>; José Reynaldo Setti<sup>2</sup>

**Resumo:** Neste artigo, propõe-se um método para produzir, por meio de um simulador, uma série de dados de tráfego sintéticos que representem diversas condições operacionais e comportamentais, como os fornecidos por sensores instalados nas vias. O método proposto baseia-se num algoritmo genético (AG), que automaticamente calibra um modelo de simulação microscópica. Entre os cromossomos (soluções) gerados pelo AG na calibração do simulador, são escolhidos os que produzem resultados marginalmente inferiores ao ótimo. As distribuições de valores dos parâmetros que formam os cromossomos deste subconjunto são usadas para gerar, através de uma amostragem de Monte Carlo, parâmetros para as replicações da simulação que são usadas para produzir os dados sintéticos. A viabilidade do método proposto é demonstrada através de um estudo de caso, no qual o simulador usado foi o TWOPAS e os dados de tráfego foram obtidos de sensores instalados no km 98 da SP98, uma rodovia de pista simples no estado de São Paulo.

**Abstract:** The paper proposes a method for generating, by a simulator, synthetic traffic data which represents a wide range of operational and behavioral conditions, such as those obtained from detectors installed on roads. The proposed method uses a genetic algorithm (GA) to calibrate a traffic simulation model automatically. A subset of chromosomes (solutions) generated by the GA during the calibration of the simulator containing those that produce results marginally inferior to the best solution is used to produce distributions of feasible values for the simulation model parameters. The distributions of simulation model parameters contained in the chromosomes of this subset of solutions are used to produce, by Monte Carlo sampling, parameters for the simulations that generate the synthetic data. The feasibility of the proposed approach is demonstrated by a case study using TWOPAS and data from detectors installed on SP98, a two-lane highway in the state of São Paulo, Brazil.

## 1. INTRODUÇÃO

Os detectores de tráfego cada vez mais vêm sendo utilizados no monitoramento do tráfego em vias urbanas e rurais, principalmente devido à facilidade de se obter dados durante todo o tempo a um custo relativamente baixo, o que não acontece com métodos “manuais” de coleta de dados do campo (Gajewski *et al.*, 2002). O uso de sensores para monitoramento de tráfego é um dos fulcros dos sistemas inteligentes de transporte, que dependem de informações constantes sobre o tráfego.

Muitas técnicas têm sido aplicadas para prever o comportamento do tráfego a partir de dados de detectores. Essas técnicas geram dados sintéticos, que podem ser velocidades ou tempos de viagem (Petty *et al.*, 1998; Lin *et al.*, 2004; Jeong e Rilett, 2005; Guo e Jin, 2006; Oh e Chung, 2006; Emam e Al-Deek, 2006) ou ainda matrizes OD (Gajewski *et al.*, 2002; Park *et al.*, 2008). Diversos estudos usam dados gerados por simulação para complementar dados coletados em campo (Schultz, 2003; Gomes, 2004) ou até para substituí-los (Araújo e Setti, 2008).

Os *softwares* de simulação de tráfego microscópicos são fortemente baseados em modelos que reproduzem características dos motoristas, como a agressividade e a escolha da rota, e dos veículos, como o modelo de desempenho. Sendo assim, diferentes resultados podem ser obtidos, dependendo do tipo de motorista/veículo que é selecionado em cada instante do período de simulação. As seleções são realizadas a partir de uma sequência de números aleatórios, que acarretam em resultados idênticos caso essa sequência não se altere. No entanto, dificilmente, uma única simulação reproduz uma situação observada em campo. A solução tradicionalmente aplicada consiste em simular uma mesma corrente de tráfego algumas vezes, mas de forma que os veículos interajam de maneiras distintas. Para isso, são realizadas replicações alterando-se somente a sequência de números aleatórios, ou, em outras palavras, mudando sua “semente”. Para fins de análise, geralmente recomenda-se o uso da média aritmética da medida de desempenho desejada, obtidas nas replicações (FHWA, 2004; Hollander e Liu, 2008; PTV, 2010).

Um exemplo da aplicação dessa abordagem tradicional consiste nos experimentos realizados para obter os modelos de tráfego propostos pelo HCM2000. Para determinar relações fluxo-velocidade em rodovias de pista simples, por exemplo, somente cinco replicações foram usadas para cada corrente de tráfego simulada, com os valores obtidos para a velocidade sendo muito próximos (Harwood *et al.*, 1999). A resposta para esse fenômeno pode estar na manutenção dos parâmetros

<sup>1</sup> José Elievam Bessa Júnior, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, Brasil. (e-mail: elievam@yahoo.com.br).

<sup>2</sup> José Reynaldo Setti, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, Brasil. (e-mail: jrasetti@usp.br).

Manuscrito recebido em 25/5/2010 e aprovado para publicação em 3/6/2010. Este artigo é parte de TRANSPORTES, volume XVIII, número 3, setembro de 2010. ISSN: 2237-1346 (online).

de calibração usados nos simuladores em cada replicação.

Muitas vezes, os usuários de simulação escolhem o número de replicações de uma corrente de tráfego de maneira arbitrária. Embora não seja um processo trivial, alguns trabalhos (FHWA, 2004; Hollander e Liu, 2008) apresentam técnicas para determinar o número mínimo de replicações de modo que a medida de desempenho obtida seja estatisticamente aceitável, próxima do valor coletado em campo. Mesmo assim, esses métodos são fundamentados em outras escolhas arbitrárias, como a determinação do desvio padrão da medida de desempenho a ser analisada, que pode ser adotado com base na experiência do usuário. Esses processos tendem a determinar o número de replicações necessário para se ter uma boa estimativa sobre a média, e não sobre a variância observada em campo. Além disso, com base na teoria do limite central, sabe-se que quanto maior o tamanho da amostra (número de replicações), mais próxima a distribuição será de uma curva normal. Ou seja, os resultados tornar-se-ão semelhantes, mesmo com a utilização de diferentes sementes de números aleatórios (FHWA, 2004).

Vários trabalhos têm demonstrado a importância de investigar a variabilidade dos resultados de simulação com base somente na mudança das sementes de números aleatórios. Lawe *et al.* (2009), usando o simulador TRANSIMS, constataram que a alteração das sementes provocou resultados com uma variabilidade muito pequena, com coeficientes de variação, em média, por volta de 0,4%. Zeeshan e Hellinga (2008) compararam os valores de atrasos em uma interseção medidos em campo com os simulados no VISSIM. Os autores também observaram que o processo de somente alterar a sequência de números aleatórios produz dados de tráfego semelhantes, o que não traduz o comportamento em campo. Isso é indesejável, uma vez que a variabilidade dos atrasos é um fator importante na análise de alternativas. Castiglione *et al.* (2003) demonstraram ainda que, em muitos casos, poucas replicações são necessárias para convergir ao valor médio de uma medida de desempenho. Em alguns casos, se a calibração do simulador for corretamente realizada, sequer são necessárias replicações.

Com base nesse contexto, o objetivo principal da pesquisa relatada neste artigo foi propor um método para produzir dados de tráfego sintéticos através de simulações que incorporem a variabilidade comportamental e operacional observada nos dados coletados por laços indutivos. A aplicação do método proposto é demonstrada através de um estudo de caso, que trata de um trecho de rodovia de pista simples no estado de São Paulo.

## 2. MÉTODO PROPOSTO

O método proposto consiste em gerar, através de um simulador calibrado para representar rodovias típicas do estado de São Paulo, um conjunto sintético de dados de tráfego semelhante aos obtidos por sensores instalados numa rodovia. Tradicionalmente, a variabilidade dos resultados de simulações é obtida variando-se as sementes das sequências de números aleatórios usadas nas simulações. O simulador é calibrado previamente e usa-se o conjunto de valores dos parâmetros de simulação que minimiza as diferenças entre o fluxo observado e o simulado (FHWA, 2004; Egami *et al.*, 2006; Araújo e Setti, 2008; Mon-Ma, 2008). Esta abordagem, no entanto, limita a variabilidade dos resultados das simulações. A inovação introduzida no método proposto consiste em introduzir variabilidade no tráfego simulado através da variação dos valores dos parâmetros de calibração do simulador. No processo tradicional, usam-se os parâmetros de calibração que minimizam as diferenças entre o fluxo de tráfego observado e o simulado. No processo proposto, os valores dos parâmetros de calibração variam de simulação para simulação e são obtidos de distribuições obtidas a partir de resultados intermediários do obtidos no processo de calibração do simulador, que é feito através de um algoritmo genético.

O fluxograma da Figura 1 resume o método proposto. O método inicia-se com a obtenção de dois conjuntos de dados. O primeiro, proveniente de um sensor selecionado, cobre diversos meses de observação e é usado para obter a distribuição de frequência das taxas de fluxo de tráfego. Os dados desse conjunto são o número de automóveis e de veículos pesados e a velocidade média observados em intervalos de 15 minutos de duração. O segundo conjunto de dados é coletado manualmente ao longo de apenas algumas horas de observação e serve para a calibração e validação do simulador. Os dados deste segundo conjunto são os tempos de viagem gastos, por cada veículo, para percorrer trechos da rodovia para a qual o simulador será calibrado, sob diversas condições de tráfego (leve ou intenso).

Os dados coletados através do sensor caracterizam o tráfego no trecho estudado. Esses dados são submetidos a um tratamento para eliminação de dados espúrios, como aqueles quando o tráfego é congestionado; ou com uma porcentagem indesejada de veículos pesados; ou mesmo aqueles obtidos quando o sensor não funcionava corretamente. Após essa etapa, dos dados restantes, obtém-se a distribuição de frequência observada das taxas de fluxos.

Os dados coletados em campo são usados para calibrar e validar o simulador, por meio de um algoritmo genético (Egami *et al.*, 2006). Do conjunto de soluções viáveis (conjunto de valores dos parâmetros de

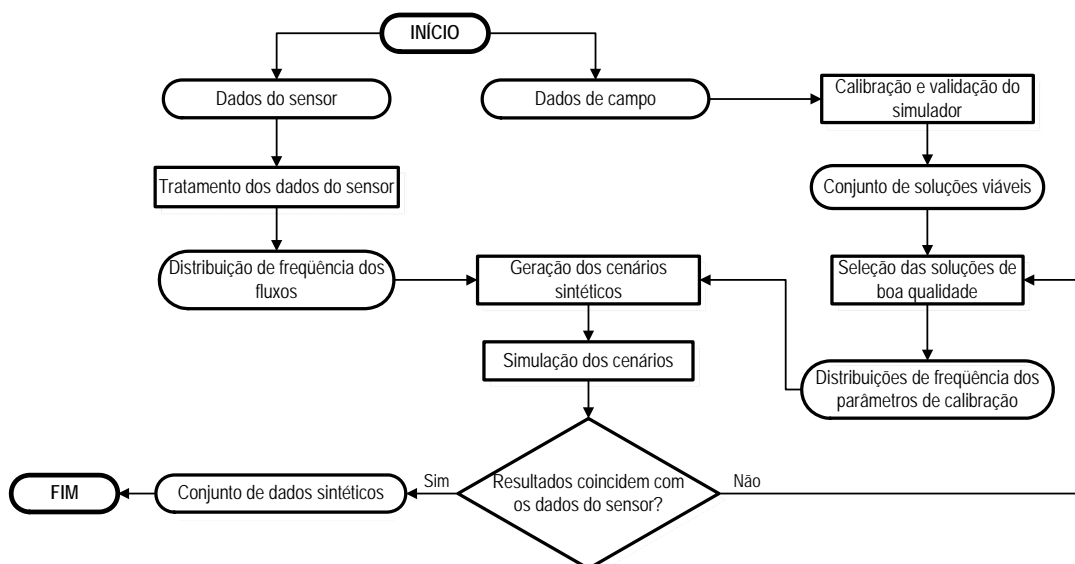


Figura 1. Fluxograma do método proposto

calibração que produzem resultados viáveis) produzidas pelo algoritmo genético durante a busca da solução ótima (a que minimiza a diferença entre o fluxo observado e o simulado), seleciona-se um subconjunto de “boas soluções” – ou seja, que produzam diferenças entre os fluxos simulado e observado inferiores a um limite previamente estabelecido. Este subconjunto de “boas soluções” é usado para introduzir variabilidade nos resultados da simulação e a partir dele são produzidas as distribuições de frequência dos valores dos parâmetros de calibração do modelo.

Um conjunto de cenários para geração de dados sintéticos é gerado na etapa seguinte. Cada cenário consiste de uma taxa de fluxo e de um conjunto de parâmetros de calibração do simulador. A seguir, os cenários são simulados e faz-se a comparação entre os resultados das simulações e os dados obtidos através do sensor. Se o conjunto de resultados for considerado representativo do tráfego na via, os resultados são os dados sintéticos; caso contrário, deve-se refinar a distribuição dos parâmetros de calibração até que o conjunto de dados sintético seja de boa qualidade.

Cada cenário sintético é criado através de uma técnica conhecida como amostragem de Monte Carlo (McLean, 1989, p. 184–185), que consiste em obter o valor de uma variável através de um número aleatório e de uma distribuição de probabilidade conhecida. O valor da variável é obtido admitindo-se que o número aleatório gerado corresponde à probabilidade acumulada. No caso, as distribuições de frequência relativa acumulada observada do fluxo bidirecional e dos parâmetros de calibração são estimadores das distribuições de probabilidade usadas para gerar os cenários sintéticos.

O simulador escolhido para gerar os dados sintéticos foi o TWOPAS, um programa para simulação de fluxos de tráfego em rodovias de pista simples disponibilizado pelo FHWA como parte do *Interactive Hi-*

*ghway Safety Design Model* (IHSDM), no qual o TWOPAS é usado no módulo para análise de tráfego (FHWA, 2007). O TWOPAS é um modelo microscópico de simulação de tráfego e foi usado na elaboração do capítulo de rodovias de pista simples do *Highway Capacity Manual* (TRB, 2000). As medidas de desempenho que o modelo é capaz de produzir incluem a velocidade média de percurso e a porcentagem do tempo viajando em pelotões. O TWOPAS é ainda capaz de incluir na simulação qualquer combinação de rampas, curvas, restrições de visibilidade, trechos de ultrapassagem proibida e faixas adicionais. O módulo de análise de tráfego do IHSDM serve para avaliar o desempenho de alternativas para a seção transversal ao longo do trecho de rodovia estudado, como faixas adicionais e pequenos segmentos com pista dupla, principalmente nas fases de preliminares do projeto (FHWA, 2007). Além de já ter sido extensivamente testado – o que garante a qualidade dos resultados obtidos –, o simulador pode ser obtido gratuitamente no site do FHWA e recebe suporte constante, o que faz dele uma alternativa atraente para uso em estudos como este.

Uma vez que os cenários sintéticos são simulados, procede-se à comparação dos dados simulados com os obtidos com o sensor. Caso a discrepância seja maior que a desejada, seleciona-se um novo conjunto de soluções geradas pelo algoritmo genético, que produzam soluções melhores que as do conjunto anterior e repetem-se os passos do procedimento até que o conjunto de dados sintéticos possa ser considerado adequado para os propósitos desejados.

As próximas seções descrevem a aplicação do método proposto usando-se dados coletados pelos detectores localizados no km 98 da SP98, uma rodovia de pista simples que liga os municípios paulistas de Bertioiga e de Mogi das Cruzes. As informações do tráfego são referentes ao ano de 2007 e foram fornecidas

pelo DER-SP.

### 3. TRATAMENTO DOS DADOS DO SENSOR

Os dados de tráfego disponibilizados pelo DER-SP cobrem o período de 1/1 a 31/12/2007 e estão agregados em 15 minutos. Os dados foram obtidos a partir de pares de laços indutivos e consistem no número de veículos que passam pelo sensor no intervalo, na porcentagem de veículos pesados (longos) e na velocidade média dos veículos que passam pelo sensor no intervalo.

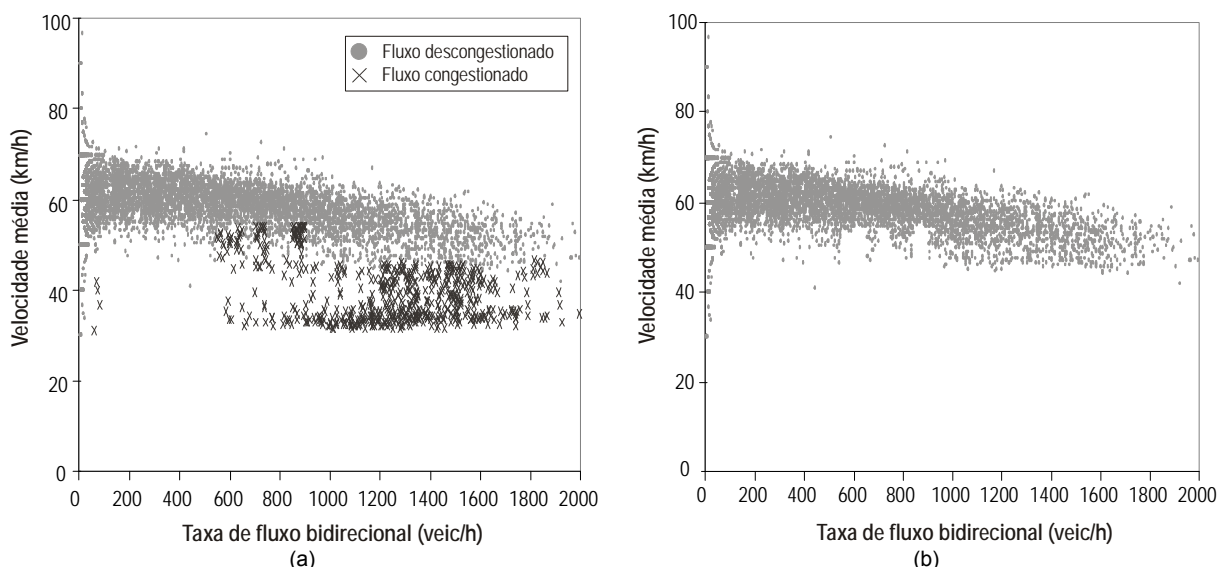
Inicialmente, foi determinada a taxa de fluxo de tráfego correspondente a cada intervalo de 15 minutos, para a elaboração da relação fluxo-velocidade do local. Como um dos objetivos da pesquisa da qual este estudo faz parte era obter uma curva fluxo-velocidade representativa de rodovias de pista simples do estado de São Paulo, foram usados apenas os intervalos sem uma porcentagem significativa de veículos longos, cujo desempenho pode ser bastante inferior ao dos automóveis. Com base numa revisão da literatura e em dados similares coletados na Rodovia dos Bandeirantes, na faixa ao lado do canteiro central, na qual o tráfego de caminhões e ônibus é proibido, decidiu-se usar apenas os intervalos nos quais se registrou a presença de até 3% de veículos longos, como mostra o gráfico da Figura 2(a). Como nesta porcentagem de veículos longos estão incluídas falsas detecções (por ex., quando dois automóveis muito próximos passam pelo sensor), a fração real de veículos pesados deve ser menor do que isso.

As observações assim selecionadas incluem tanto intervalos em que o fluxo é congestionado como intervalos em que o fluxo é descongestionado. A segunda etapa do tratamento dos dados consiste separar as observações relativas a condições de fluxo desconges-

tionado. Para isso, foi usada análise de *cluster*. As observações foram agrupadas em faixas de fluxo de 50 veic/h de largura: 0–50, 50–100 veic/h, etc. Apenas as faixas em que existiam regiões com fluxo descongestionado e congestionado foram selecionadas para a análise de *cluster*, já que não seria preciso aplicar o processo àquelas em que apenas há fluxo descongestionado.

A velocidade foi definida como critério de agrupamento e a distância euclidiana foi escolhida para medir a similaridade entre os objetos. O algoritmo de aglomeração adotado foi o método não-hierárquico *k-means*, recomendado para grandes bases de dados, como é o caso abordado (Hair Jr. *et al.*, 2005). A Figura 2(a) mostra o resultado da análise de *cluster*, indicando as observações em cada regime de fluxo, enquanto que a Figura 2(b) apresenta apenas as observações referentes a condições de fluxo descongestionado, que foram utilizadas neste estudo de caso.

Os dados usados para calibrar e validar o TWOPAS foram obtidos em duas coletas (com 4 horas cada), realizadas com câmeras filmadoras posicionadas nas extremidades do trecho que inclui o detector, entre o km 93,8 e o km 97,8, nos dias 23/11 e 7/12/2008. Dessa forma, foram determinados a velocidade média de percurso (que corresponde à velocidade média no espaço) e a taxa de fluxo de veículos para intervalos de 15 minutos de duração. Este trecho pode ser considerado de condições ideais: localiza-se em região de relevo plano, tem alinhamento horizontal tangente, com apenas uma curva suave próxima à metade, possui acostamento pavimentado, faixas de tráfego de 3,50 m de largura e não há faixas adicionais. Nas duas coletas, o tráfego era composto praticamente apenas por automóveis (96,5% na primeira coleta e 97% na segunda). As datas e o local da coleta foram escolhidos



**Figura 2.** Tratamento dos dados de tráfego do sensor da SP 98 (km 98): (a) observações com até 3% de veículos longos, nos regimes de fluxo descongestionado e fluxo congestionado; (b) observações com até 3% de veículos longos, regime de fluxo descongestionado

para que fosse possível observar a corrente de tráfego sob grande demanda, uma vez que se desejava calibrar o TWOPAS com dados de fluxo próximos à capacidade. Nas duas coletas, em média, foram observados fluxos bidirecionais variando entre, aproximadamente, 1000 e 1650 veic/h e distribuições direcionais variando entre 95%/5% e 82%/18%.

Conforme mostra a Figura 3(a), os pontos fluxo-velocidade dos dois conjuntos de dados coletados (agregados a cada 15 minutos) apresentaram uma diferença na distribuição em relação aos pontos do laço indutivo. Esta diferença é explicada pela localização do sensor, que fica ao lado de um posto da Polícia Rodoviária, onde os motoristas reduzem voluntariamente a velocidade por medo da fiscalização.

Para correção deste efeito, foi realizado um tratamento simplificado dos dados do sensor para o regime descongestionado: foi calculada a diferença média entre as velocidades observadas em campo e aquelas determinadas, para uma mesma taxa de fluxo, de uma função linear obtida para os dados do sensor, cuja forma é  $v = 77 - 0,00524 \cdot q$ , em que  $v$  representa a ve-

locidade média de viagem e  $q$  é o fluxo de tráfego bidirecional. Esta diferença média, de 15 km/h, foi, então, somada a cada velocidade medida pelo laço indutivo, tornando-as próximas das velocidades médias observadas na coleta de dados. O resultado do deslocamento pode ser visto na Figura 3(b).

Percebe-se que, dessa forma, a velocidade de fluxo livre tornou-se próxima de 80 km/h, compatível com o limite máximo permitido no trecho. Um teste de Kolmogorov-Smirnov ( $\alpha = 0,05$ ) sugere que as velocidades médias assim obtidas distribuem-se segundo uma distribuição normal, como mostra a Figura 4. No gráfico, a distribuição de frequência relativa das velocidades médias e a curva normal correspondente apresentaram-se próximas, o que confirma o resultado do teste de aderência realizado.

#### 4. CALIBRAÇÃO E VALIDAÇÃO DO SIMULADOR TWOPAS

A crescente complexidade dos projetos rodoviários e do tráfego tem influenciado a criação de métodos de análise que sejam mais flexíveis e robustos do que as

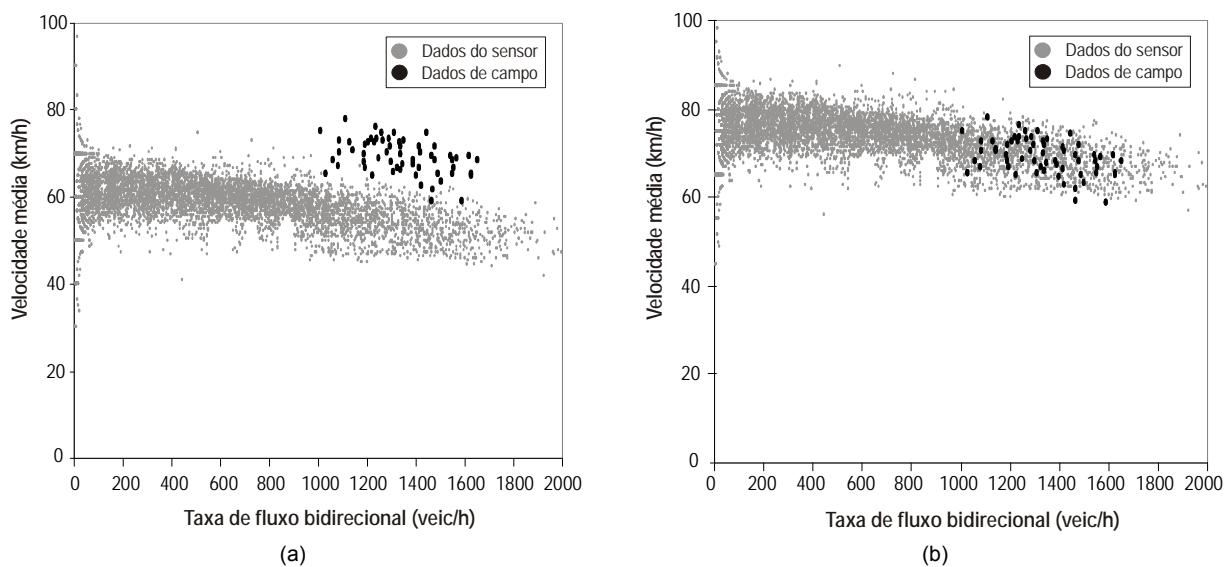


Figura 3. Tratamento dos dados do detector da SP 98 (km 98): (a) efeito da redução de velocidade por causa do posto da Polícia Rodoviária; e (b) correção desse efeito

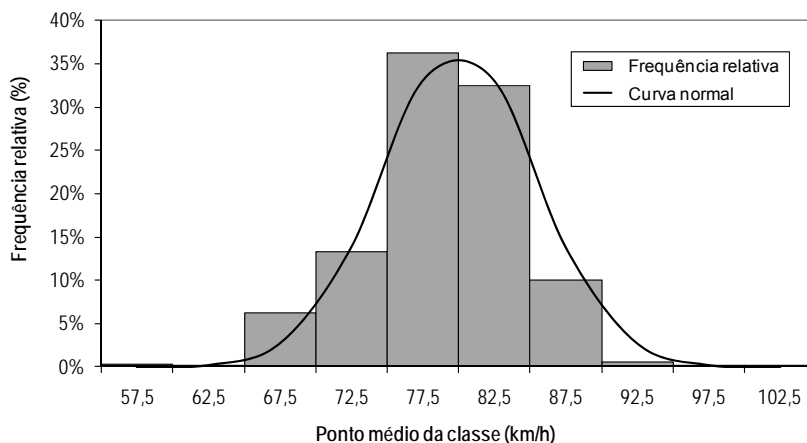


Figura 4. Distribuição de frequência das velocidades médias observadas pelo sensor após o tratamento dos dados

abordagens tradicionais, como os métodos analíticos (Hellings, 1998). Essa conjuntura, combinada com a crescente capacidade dos computadores pessoais, tem criado um ambiente propício ao desenvolvimento dos simuladores de tráfego, que podem ser utilizados para auxiliar na análise e no projeto de sistemas de transporte (McLean, 1989; Hellings, 1998). Dentre as vantagens da simulação, é possível avaliar um conjunto de alternativas sem haver a necessidade de intervir diretamente no tráfego, além de possibilitar estudar situações ainda inexistentes (Kim e Rilett, 2001).

Geralmente, um simulador de tráfego é composto por vários submodelos, cada qual incluindo vários parâmetros que devem ser calibrados para melhor representar o sistema sendo simulado. No entanto, muitos desses parâmetros são difíceis de medir em campo, seja porque são complicados de isolar, seja devido à necessidade de coletar muitos dados (Hollander e Liu, 2008). Por conveniência, parâmetros *default* são disponibilizados ao usuário, mas devem ser apenas usados para modelar locais onde as condições sejam semelhantes àquelas onde foram obtidos (FHWA, 2004). Calibrar os parâmetros de um simulador pode ser considerado um problema de otimização em que um conjunto de parâmetros deve ser determinado para maximizar (ou minimizar) uma função objetivo, que expressa a diferença entre características do tráfego observados (como velocidade, volume e tempo de viagem) e os resultados da simulação (Kim e Rilett, 2001). Uma ferramenta de otimização que vem sendo bastante utilizada para calibrar simuladores de tráfego é o algoritmo genético (Kim e Rilett, 2001; Egami *et al.*, 2004).

Os algoritmos genéticos (AGs) fazem parte do conjunto dos algoritmos evolutivos, técnicas de busca e otimização inspiradas no mundo biológico que se baseiam na teoria da evolução de Darwin (Goldberg, 1989). São técnicas robustas, que buscam soluções a partir de múltiplos pontos, sendo recomendados quan-

do o espaço de busca é complexo, grande e pouco conhecido (Kim e Rilett, 2001).

O AG é um procedimento iterativo que se inicia com a produção aleatória de uma população, um conjunto de cromossomos (indivíduos), que são soluções para o problema; esses cromossomos geralmente são representados por *strings* binárias, em que suas partes são compostas pelos parâmetros do simulador que se deseja calibrar. Durante cada iteração, também chamada de geração, os indivíduos são avaliados e selecionados segundo uma função *fitness* que mede seu grau de adaptação ao meio ambiente, ou seja, a qualidade da solução que fornecem. Isso faz com que indivíduos mais bem adaptados (que fornecem melhores soluções) permaneçam na população transmitindo as suas características ao longo das gerações, enquanto que os menos adaptados tendem a desaparecer. A utilização de AGs não garante uma convergência, o que faz com que o número máximo de gerações para encontrar um ótimo global seja impossível de ser determinado (Goldberg, 1989).

Tradicionalmente, três operadores genéticos são usados para formar uma nova geração da população: *crossover*, seleção e mutação. O primeiro usa dois indivíduos (pais) para produzir novos indivíduos (filhos) a partir da troca de partes específicas dos pais. O elitismo é o processo de seleção mais comum, que consiste em escolher o melhor indivíduo para se reproduzir com outros da população. A mutação serve para introduzir variação na solução e, por consequência, aumentar a probabilidade de se atingir um ótimo global (Goldberg, 1989).

#### 4.1. Implementação do programa para calibração automática do TWOPAS

O programa para calibração automática do TWOPAS usado foi implementado em Perl e foi baseado numa versão usada em um estudo anterior (Egami *et al.*, 2006; Mon-Ma, 2008). A Figura 5 mostra o funcio-

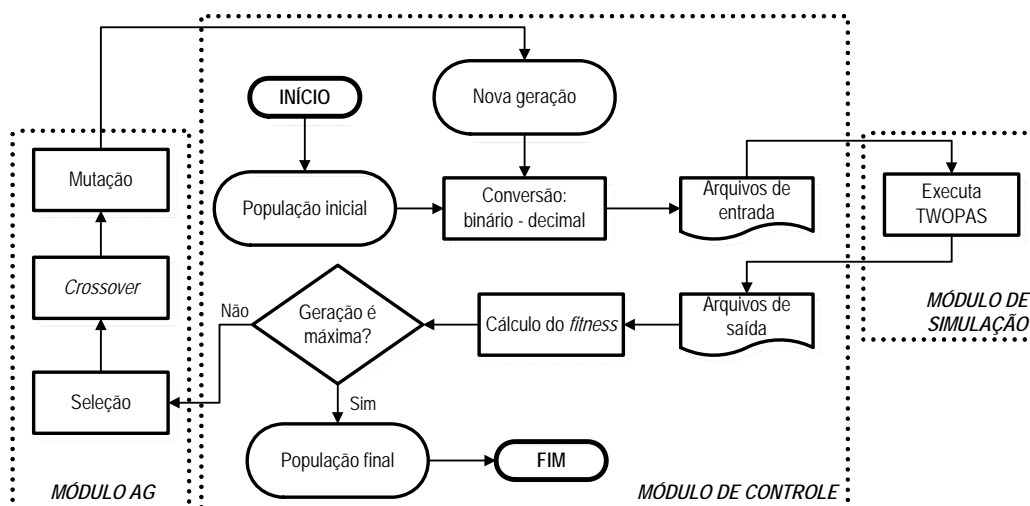


Figura 5. Fluxograma do programa utilizado na calibração automática do TWOPAS

namento do programa, que é composto de três módulos: controle, simulação e AG. O módulo de controle contém rotinas para desenvolvimento de uma das principais vantagens do AG: a automatização do processo de calibração. Essas rotinas incluem a criação da população inicial de 20 cromossomos; a conversão dos cromossomos de *strings* binárias para valores decimais dos parâmetros de calibração e vice-versa; a inserção dos parâmetros nos arquivos de entrada do TWOPAS; a extração de medidas de desempenho dos arquivos de saída do simulador e cálculo da função *fitness*; e a verificação do critério de parada, que é o número de gerações (130).

Após a produção do arquivo de entrada, o módulo de simulação aciona a execução do TWOPAS. O tempo de simulação adotado foi de 60 minutos, com *warm-up* de 10 minutos. Para aumentar a confiabilidade da calibração, as correntes de tráfego foram replicadas cinco vezes para diferentes conjuntos de sementes de números aleatórios, para cada cenário sintético. Em caso do critério de parada não ter sido atingido, o módulo do AG é executado e os operadores genéticos (seleção, *crossover* e mutação) produzem uma nova geração da população. O elitismo foi o critério de seleção adotado; a taxa de *crossover* adotada foi 0,5 e a de mutação, 0,3. O processo se reinicia até o critério de parada ser atingido. A população final fornece os valores dos parâmetros de calibração que produzem a menor diferença entre o fluxo observado na rodovia e o simulado pelo TWOPAS.

Os cromossomos (indivíduos) usados no AG são formados pelos parâmetros de calibração que mais a-

fetam as medidas de desempenho adotadas, que são a velocidade média de percurso e a porcentagem de tempo viajando em pelotões. Esses parâmetros foram determinados através de uma análise de sensibilidade realizada num estudo anterior (Mon-Ma, 2008). Na Tabela 1, são mostrados as definições e os intervalos de busca dos parâmetros de calibração, além da quantidade de *bits* que cada parâmetro ocupa no cromossomo. A *string* que corresponde ao cromossomo tem 59 *bits*.

Os parâmetros  $BKPM(j)$  e  $WOHP(i)$  são calculados através de outras variáveis, como mostram as Equações 1 e 2:

$$BKPM_{cal}(j) = BKPM_{def}(j) + \Delta comp, \quad j = 1, \dots, 10, \quad (1)$$

em que  $BKPM_{cal}(j)$  é o fator estocástico do tipo de motorista calibrado para o motorista  $j$  e  $BKPM_{def}(j)$  é o fator estocástico do tipo de motorista *default* para o motorista  $j$ . A variável  $\Delta comp$  também deve ser a mesma para determinar cada um dos 10 parâmetros  $BKPM$ . A relação massa/potência dos caminhões da classe  $i$  é calculada por:

$$WOHP_{cal}(i) = \frac{1}{\eta} \left( \frac{massa_{real}(i)}{P_{nom}(i)} \right), \quad i = 1, \dots, 4, \quad (2)$$

em que,

$WOHP_{cal}(i)$ : relação massa/potência calibrada para o caminhão tipo  $i$ ;

$massa_{real}(i)$ : massa do caminhão tipo  $i$ ;

$P_{nom}(i)$ : potência nominal do caminhão tipo  $i$ .

**Tabela 1.** Definições, intervalos de busca e tamanho da string dos parâmetros de calibração do TWOPAS

Parâmetro	Definição	Intervalo de busca		
		Mínimo	Máximo	Bits
<i>PREC</i>	probabilidade de o motorista reconsiderar uma ultrapassagem no período de revisão	0,1	0,5	3
<i>VEAN</i>	velocidade média desejada de todos os motoristas	82 km/h	104 km/h	5
<i>VSIG(1,1)</i>	desvio padrão da velocidade média desejada na direção 1 da categoria veicular 1 (caminhões)	5 km/h	20 km/h	4
<i>VSIG(1,3)</i>	desvio padrão da velocidade média desejada na direção 1 da categoria veicular 3 (automóveis)	5 km/h	20 km/h	4
<i>VSIG(2,1)</i>	desvio padrão da velocidade média desejada na direção 2 da categoria veicular 1	5 km/h	20 km/h	4
<i>VSIG(2,3)</i>	desvio padrão da velocidade média desejada na direção 2 da categoria veicular 3	5 km/h	20 km/h	4
<i>VBI(1,1)</i>	diferença entre <i>VEAN</i> e a velocidade média desejada na direção 1 da categoria veicular 1	-22 km/h	5 km/h	5
<i>VBI(1,2)</i>	diferença entre <i>VEAN</i> e a velocidade média desejada na direção 1 da categoria veicular 2	-5 km/h	22 km/h	5
<i>VBI(2,1)</i>	diferença entre <i>VEAN</i> e a velocidade média desejada na direção 2 da categoria veicular 1	-22 km/h	5 km/h	5
<i>VBI(2,3)</i>	diferença entre <i>VEAN</i> e a velocidade média desejada na direção 2 da categoria veicular 3	-5 km/h	22 km/h	5
<i>ZKCOR</i>	fator de sensibilidade do modelo de <i>car-following</i>	0,6	1,0	3
$\Delta comp$	usada para calcular $BKPM(k)$ , fator estocástico do motorista tipo $k$ , $k = 1, \dots, 10$	-0,20	0,20	6
$\eta$	usada no cálculo da relação massa/potência $WOHP(i)$ , do caminhão tipo $i$ , $i = 1, \dots, 4$	0,60	1,20	6

Os dados da massa e da potência nominal dos caminhões utilizados na calibração foram obtidos num estudo anterior (Cunha *et al.*, 2005), no qual as massas foram coletadas em balanças rodoviárias, enquanto que a potência nominal em catálogos de revistas especializadas. A relação massa/potência real foi obtida para cada categoria de caminhões modelada no TWOPAS e o valor encontrado pelo AG para  $\eta$  foi admitido como o mesmo para as quatro categorias.

A função *fitness* adotada para avaliar o grau de adaptação dos indivíduos é a mesma usada num estudo similar (Mon-Ma, 2008):

$$fitness = 100 \cdot e^{-5MAER}, \quad (3)$$

em que *MAER* é a razão do erro médio absoluto (*mean absolute error ratio*) das medidas de desempenho usadas.

O *MAER* representa a diferença entre as medidas de desempenho observadas em campo e as simuladas. A função *MAER* usada leva em consideração os valores médios das medidas de desempenho adotadas (velocidade média de percurso e porcentagem de veículos em pelotões), além dos valores do 15° e do 85° percentis da distribuição de velocidades médias de percurso. Esses percentis foram incluídos na tentativa de considerar a dispersão da distribuição de velocidades, e não apenas valores médios. A formulação adotada para *MAER* é:

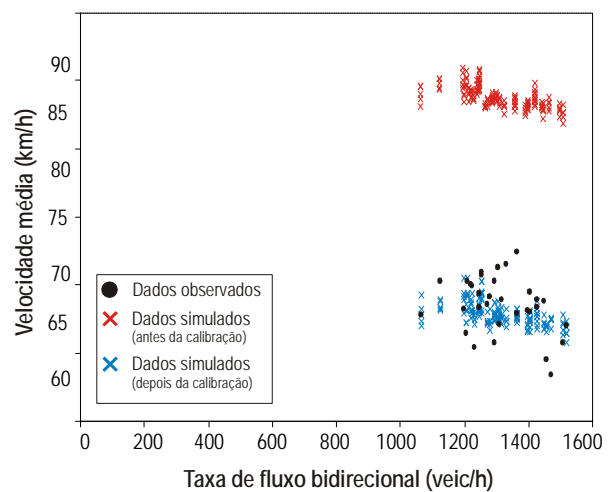
$$MAER = \frac{1}{14} \cdot \left( \sum_{i=1}^2 \sum_{j=1}^2 \left| \frac{V_{obs}(i,j) - V_{sim}(i,j)}{V_{obs}(i,j)} \right| + \sum_{i=1}^2 \left| \frac{PF_{obs}(i) - PF_{sim}(i)}{PF_{obs}(i)} \right| + \sum_{i=1}^2 \sum_{j=1}^2 \left| \frac{V_{15obs}(i,j) - V_{15sim}(i,j)}{V_{15obs}(i,j)} \right| + \sum_{i=1}^2 \sum_{j=1}^2 \left| \frac{V_{85obs}(i,j) - V_{85sim}(i,j)}{V_{85obs}(i,j)} \right| \right), \quad (4)$$

em que,

- $V_{sim}(i,j)$ : velocidade média de percurso simulada da categoria veicular  $j$  (automóveis ou caminhões) na direção  $i$ ;
- $V_{obs}(i,j)$ : velocidade média de percurso observada da categoria veicular  $j$  na direção  $i$ ;
- $PF_{sim}(i)$ : porcentagem de veículos em pelotões simulada na direção  $i$ ;
- $PF_{obs}(i)$ : porcentagem de veículos em pelotões observada na direção  $i$ ;
- $V_{15sim}(i,j)$ : 15° percentil da distribuição de velocidades de percurso simuladas da categoria veicular  $j$  na direção  $i$ ;
- $V_{15obs}(i,j)$ : 15° percentil da distribuição de velocidades de percurso observadas da categoria veicular  $j$  na direção  $i$ ;

- $V_{85sim}(i,j)$ : 85° percentil da distribuição de velocidades de percurso simuladas da categoria veicular  $j$  na direção  $i$ ;
- $V_{85obs}(i,j)$ : 85° percentil da distribuição de velocidades de percurso observadas da categoria veicular  $j$  na direção  $i$ ;

Finalizado o processo de calibração, o simulador recalibrado foi submetido a uma validação, que consistiu na comparação, através de gráficos, dos valores das medidas de desempenho observadas e simuladas. A Figura 6 mostra as diferenças entre as velocidades médias de percurso observadas em campo (nas duas coletas) e as simuladas antes e depois da calibração, respectivamente. É possível perceber a eficiência da calibração comparando-se o resultado das simulações com as observações feitas na rodovia.



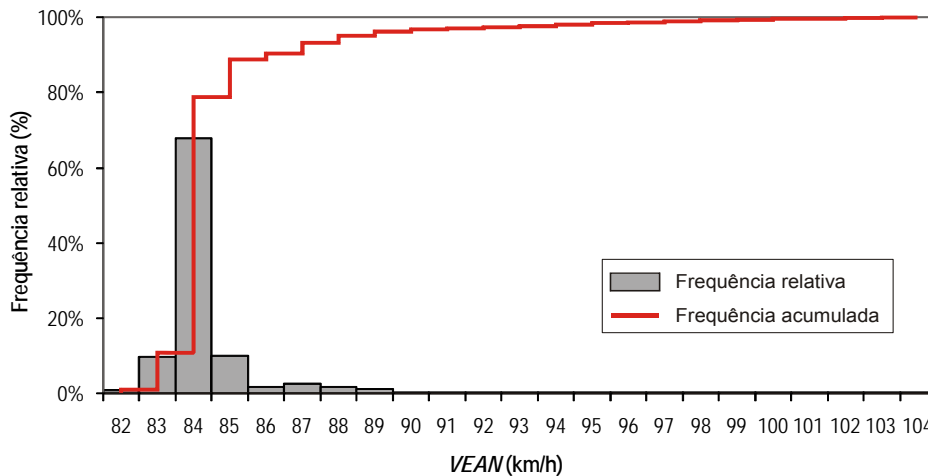
**Figura 6.** Velocidades médias de percurso observadas e simuladas em função da taxa de fluxo de tráfego bidirecional, antes da aplicação do AG, usando-se os valores default dos parâmetros de calibração, e com o modelo recalibrado pelo AG

## 4.2. Seleção das soluções de boa qualidade

Ao contrário da maioria dos trabalhos que tratam de calibração de simuladores, esta pesquisa procurou obter os parâmetros de calibração através de distribuições de frequência, ao invés de apenas considerar um valor único para cada parâmetro. A Figura 7 mostra a distribuição de frequência do parâmetro de calibração *VEAN*, a velocidade média desejada por todos os motoristas. Pode-se notar que o valor mais frequente nos cromossomos testados foi 84 km/h, o que indica que este valor provavelmente é o ideal, mas valores de até 104 km/h foram também testados em uns poucos casos.

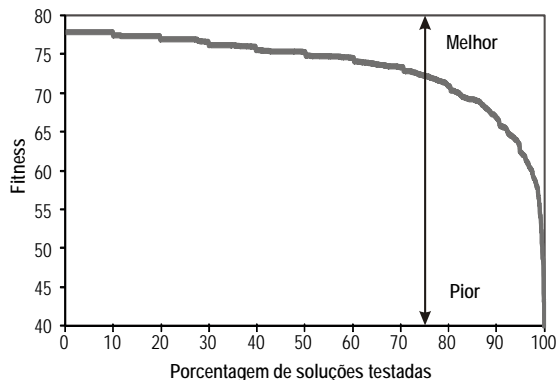
Como o objetivo do estudo é gerar dados sintéticos com variabilidade similar à observada na rodovia, a abordagem adotada requer a obtenção de um subconjunto de soluções que contenha valores dos parâmetros de calibração que forneçam resultados com erro inferior a um patamar previamente estabelecido. O cri-





**Figura 7.** Distribuição de frequência dos valores encontrados para VEAN, considerando o conjunto de todas as soluções testadas pelo AG

tério de seleção adotado baseou-se no valor da função do *fitness* das soluções testadas pelo AG ao longo das 130 gerações. A Figura 8 mostra que, à medida que o *fitness* aumenta, o tamanho do conjunto de soluções testadas diminui. O maior *fitness* obtido fica em torno de 78 (100 é o máximo) e o menor *fitness* é inferior a 50. O *fitness* que limita 50% das soluções é igual a 75.



**Figura 8.** Variação do *fitness* em função da porcentagem de soluções testadas na calibração

Como é característica em calibrações utilizando AG, uma grande parte dos cromossomos testados ao longo das gerações possui *fitness* com valor alto, próximo ao da melhor solução encontrada. Esse fenômeno ocorre por duas razões: (i) as características da função objetivo e do próprio simulador, que possui um grande número de submodelos e de parâmetros de calibração; e (ii) como a otimização realizada pelo AG é bastante eficiente, as piores soluções são logo excluídas da população. No entanto, como o algoritmo continua atuando durante mais gerações para procurar um ótimo global, uma grande quantidade de cromossomos que produz resultados apenas marginalmente piores que o ótimo é criada ao longo do processo.

A escolha do limite que define o subconjunto das melhores soluções foi realizada através de uma análise de sensibilidade, que é apresentada no item a seguir, juntamente com a produção da série de dados de tráfego sintéticos.

## 5. PRODUÇÃO DOS DADOS DE TRÁFEGO SINTÉTICOS

A principal característica do método proposto consiste em selecionar cromossomos com *fitness* maior ou igual a um dado limite. Com este conjunto de soluções, obtém-se a distribuição de valores para cada um dos parâmetros de calibração do simulador, similar à mostrada na Figura 7. A geração dos dados sintéticos é feita através dos seguintes passos:

Determinar o número de observações desejado  $n$ ;

Para  $i = 1$  a  $n$ ,

Gerar a taxa de fluxo  $V(i)$  a partir da distribuição acumulada de frequência da taxa de fluxo;

Gerar os valores dos parâmetros de calibração  $C(i)$ , a partir das distribuições acumuladas de frequência dos parâmetros de calibração;

Escolher as sementes dos números aleatórios para as cinco replicações da simulação;

Simular a corrente de tráfego usando a taxa de fluxo  $V(i)$ , os parâmetros  $C(i)$  e as sementes dos números aleatórios;

Obter a taxa de fluxo e a velocidade média da corrente de tráfego simulada a partir dos resultados das replicações da simulação;

Repetem-se os passos anteriores até  $i = n$ ;

Os primeiros três passos (gerar a taxa de fluxo de tráfego  $V(i)$ , gerar o vetor  $C(i)$  de valores dos parâmetros de calibração e escolher as sementes de números aleatórios para as cinco replicações de cada simulação) constituem-se na criação do cenário sintético. Os demais passos produzem uma das  $n$  observações que formam o conjunto de dados sintéticos que se deseja produzir.

A geração da taxa de fluxo e dos valores dos parâmetros de calibração é feita através de um processo de

amostragem de Monte Carlo, usando-se um número aleatório  $r$  ( $0 \leq r \leq 1$ ) e uma distribuição de frequência acumulada como a mostrada na Figura 7. Por exemplo, se o número aleatório  $r$  é 0,70, como  $r$  corresponde à frequência acumulada, VEAN deve ser igual a 84 km/h.

A definição das distribuições dos valores dos parâmetros de calibração depende do valor mínimo do *fitness* adotado na seleção dos cromossomos do subconjunto de soluções de boa qualidade. Como não há, *a priori*, uma forma de estabelecer este valor mínimo, foi realizada uma análise de sensibilidade testando-se quatro *fitness* que delimitam as seguintes porcentagens de soluções utilizadas: 15% (*fitness*  $\geq 77$ ), 50% (*fitness*  $\geq 75$ ), 85% (*fitness*  $\geq 69$ ) e 100% (*fitness*  $\geq 43$ ). Os resultados da análise de sensibilidade estão mostrados na Figura 9.

Durante a calibração, o maior *fitness* encontrado foi de 77,9, o que significa que o conjunto de 15%, 50%, 85% e 100% dos cromossomos testados delimitam o uso das soluções de até 1%, 3%, 11% e 45% inferiores ao *fitness* máximo, respectivamente, como pode ser visto na Figura 8.

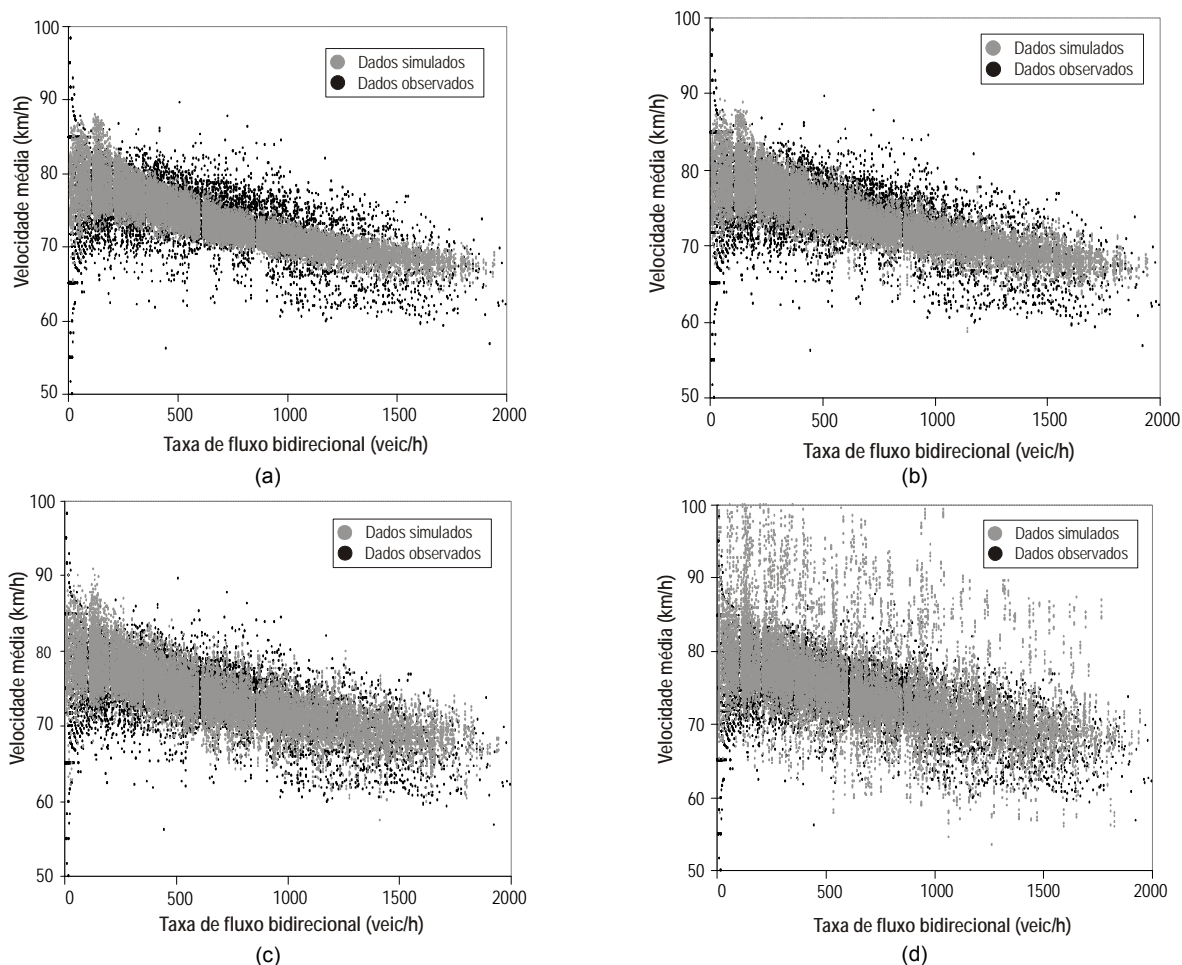
As Figuras 9(a) e 9(d) mostram os resultados obtidos considerando 15% e 100% das soluções, respectivamente.

vamente. É possível notar que, no primeiro caso, a distribuição de pontos fluxo-velocidade obtida apresenta uma variabilidade bem menor que a observada nos dados obtidos pelo sensor. O fenômeno inverso ocorreu com o uso de todas as soluções geradas na calibração, em consequência da alta variabilidade dos parâmetros de calibração. Nesse caso, percebe-se que a dispersão dos dados fluxo-velocidade simulados foi bem maior do que a observada para os dados provenientes do laço detector. Os gráficos mostrados nas Figuras 9(b) e 9(c) ilustram a utilização de 50% e de 85% das soluções, respectivamente.

Visualmente, percebe-se que o conjunto com 85% dos cromossomos forneceu a distribuição dos pontos fluxo-velocidade simulados mais próximos dos pontos obtidos pelo laço detector, o que pode ser considerado indicativo da qualidade das soluções geradas pelo AG usado.

## 6. CONSIDERAÇÕES FINAIS

O gráfico da Figura 10 compara os dados sintéticos gerados pelo método proposto com a prática normal de se obter variabilidade dos resultados da simulação apenas variando-se as sementes das sequências de números aleatórios usadas pelo simulador. Na Figura



**Figura 9.** Análise de sensibilidade para determinação do conjunto de boas soluções: (a) 15% das soluções (*fitness*  $\geq 77$ ); (b) 50% das soluções (*fitness*  $\geq 75$ ); (c) 85% das soluções (*fitness*  $\geq 69$ ); e (d) 100% das soluções (*fitness*  $\geq 43$ )

10(b), os pontos foram gerados usando-se os mesmos experimentos adotados na obtenção da curva fluxo-velocidade do HCM2000 (no exemplo, para velocidade de fluxo livre igual a 70 km/h), relatados em Harwood *et al.* (1999). Nesse caso, foram simulados 16 taxas de fluxo bidirecionais, entre 200 veic/h e 3200 veic/h, em intervalos de 200 veic/h, com 5 sementes diferentes das sequências de números aleatórios. A divisão direcional adotada foi de 50/50. Pode-se notar facilmente que a aplicação do método proposto possibilita gerar, através de simulação, dados de tráfego sintéticos com variabilidade comportamental e operacional semelhante à observada em dados coletados por sensores instalados numa rodovia.

O uso do método proposto pode expandir o uso de dados coletados por sensores, uma vez que os experimentos de simulação podem ser realizados em condições controladas pelo usuário e fornecem mais informações sobre as viagens. O método proposto representa um avanço na calibração de modelos de simulação, uma vez que permite controlar o grau de variabilidade dos resultados das simulações, de forma a reproduzir a variabilidade natural dos fluxos de tráfego. Apesar de o estudo de caso ter sido realizado para uma rodovia de pista simples, o método pode ser facilmente adaptado para rodovias de pista dupla.

O método proposto requer o uso de um AG para a calibração do modelo e geração do subconjunto de soluções de boa qualidade. Os resultados indicam que soluções que produzam *fitness* até cerca de 10% inferior ao melhor *fitness* observado podem ser usadas para a geração dos dados sintéticos.

Pretende-se aplicar o método para obter modelos para a relação fluxo-velocidade de rodovias de pista simples no Brasil. Geralmente, os modelos obtidos através de simuladores somente se valem de alguns pontos médios e não de uma série de dados, como apresentado neste artigo. Com a aplicação do método, as relações obtidas ganhariam em confiabilidade.

## AGRADECIMENTOS

Os autores agradecem o apoio do DER-SP que, através de um convênio de cooperação técnico-científica, forneceu dados e suporte para as coletas de dados realizadas no desenvolvimento desta pesquisa. Os autores também agradecem ao CNPq, pelo suporte financeiro sob a forma de bolsas de mestrado e de produtividade em pesquisa, e aos avaliadores deste artigo, pelas sugestões apresentadas para o seu aprimoramento.

## REFERÊNCIAS

- Araújo, J. J. e J. R. Setti (2008) Análise do efeito do tráfego de veículos pesados sobre uma ponte através de simulação microscópica. In: Transporte em Transformação XII: Trabalhos vencedores do Prêmio CNT de Produção Acadêmica 2007. CNT/ANPET, p. 23–42.
- Castiglioni, J.; J. Freedman e M. Bradley (2003) Systematic investigation of variability due to random simulation error in an activity-based microsimulation forecasting model. *Transportation Research Record* v. 1831, p.76–88.
- Cunha, A. L. B. N.; M. L. Mon-Ma; J. J. Araújo; C. Y. Egami e J. R. Setti (2005) Caracterização da frota de caminhões no sistema Anhanguera - Bandeirantes. *XIX ANPET Congresso de Pesquisa e Ensino em Transportes*, Recife-PE, 2005, CD-ROM.
- Egami, C. Y.; J. R. Setti e L. R. Rilett (2004) Algoritmo genético para calibração automática de um modelo de simulação de tráfego em rodovias de pista simples. *Transportes*, v. 12, p. 5–14.
- Egami, C.Y.; M. L. Mon-Ma; J. R. Setti e L. R. Rilett (2006) [falta colocar a referência]
- Emam, E. B. e H. Al-Deek (2006) Using real-life dual-loop detector data to develop new methodology for estimating freeway travel time reliability. *Transportation Research Record* v. 1959, p. 140–150.
- FHWA (2004) *Traffic Analysis Toolbox Volume III: guidelines for applying traffic microsimulation modeling software*. Federal Highway Administration (FHWA).
- FHWA (2007) *Traffic Analysis Module (TAM) Engineer's Manual – Interactive Highway Safety Design Model (IHSDM)*. Federal Highway Administration (FHWA), disponível em <<http://www.fhwa.dot.gov/research/tfhr/projects/safety/comprehensive/ihsdm/tamweb.cfm>>.
- Gajewski, B. J.; L. R. Rilett; M. P. Dixon e C. H. Spilgeman (2002) Robust Estimation of Origin-Destination Matrices. *Journal of Transportation and Statistics*, v. 5, p.37–56
- Goldberg, D. E. (1989) *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley Publication Company, Inc., Reading, Massachusetts.
- Gomes, G. Z. (2004) Uso de microsimulação na avaliação da sustentabilidade de corredores rodoviários. Dissertação (mestrado). Programa de Pós-Graduação em Engenharia de Transportes, EESC-USP.
- Guo, H e J. Jin (2006) Travel time estimation with correlation analysis of single loop detector data. *Transportation Research Record* v. 1968, p. 10–19.
- Hair Jr., J. F.; R. E. Anderson; R. L. Tatham e W. C. Black (2005) *Análise multivariada de dados*. 5ª Ed. Bookman Companhia Editora. Porto Alegre, RS, 593 p.
- Harwood, D. W.; A. D. May; I. B. Anderson; L. Leiman e A. R. Archilla

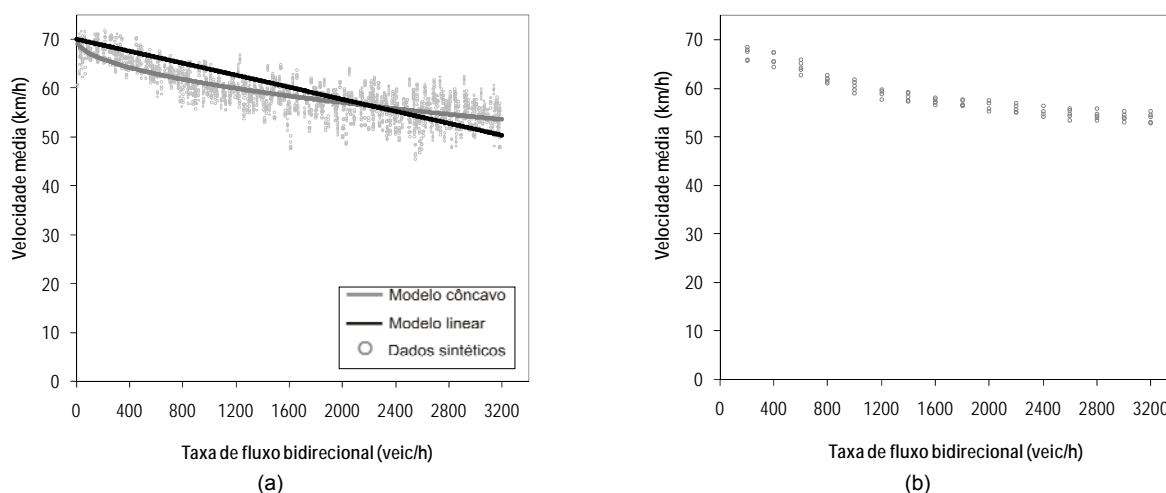


Figura 10. Comparação entre: (a) dados sintéticos obtidos com o método proposto e (b) com a abordagem tradicional

- (1999) *Capacity and quality of service of two-lane highways*. Final Report, NCHRP Project 3-55 (3). MRI, Kansas City, Mo.
- Hellinga, B. (1998) Requirements for the validation and calibration of traffic simulation models. *Proceedings of Canadian Society for civil engineering 1998*. Halifax, Nova Scotia, v. IVb, p.211–222.
- Hollander, Y. e R. Liu (2008) The principles of calibrating traffic microsimulation models. *Transportation*, v. 35, p. 347–362.
- Jeong, R. e L. R. Rilett (2005) Prediction model of bus arrival time for real-time applications. *Transportation Research Record* v. 1927, p. 195–204.
- Kim, K. e L. R. Rilett (2001) Genetic algorithm based approach for calibration microscopic simulation models. *IEEE Intelligent Transportation Systems Conference Proceedings*, Oakland, CA, USA, p. 698–704.
- Lawe, S.; J. Lobb; A. W. Sadek; S. Huang e C. Xie (2009) TRANSIMS Implementation in Chittenden County, Vermont: development, calibration, and preliminary sensitivity analysis. *Transportation Research Record* v. 2132, p.113–121.
- Leiman, L.; A. R. Archilla e A. May (1998) *Capacity and quality of service of two-lane highways*. NCHRP Project 3-55 (3). Task 6 – Enhance, calibrate, and validate the selected simulation model, TWOPAS model improvements. Midwest Research Institute, Kansas City, Mo.
- Lin, W.; J. Dahlgren e H. Huo (2004) Enhancement of Vehicle Speed Estimation with Single Loop Detectors. *Transportation Research Record* v. 1870, p.147–152.
- McLean, J. R. (1989) *Two-lane highway traffic operations – Theory and practice*. Gordon and Breach Science Publishers, Amsterdam.
- Mon-Ma, M. L. (2008) *Adaptação do HCM-2000 para rodovias de pista simples com faixas adicionais típicas do Estado de São Paulo*. 162 p. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- Oh, J e Y. Chung (2006) Calculation of travel time variability from loop detector data. *Transportation Research Record* v. 1945, p. 12–23.
- Park, E. S.; L. R. Rilett e C. H. Spielgelman (2008) A Markov chain Monte Carlo-based origin destination matrix estimator that is robust to imperfect intelligent transportation systems data. *Journal of Intelligent Transportation Systems*, v. 12, n. 3, p.139–155.
- Petty *et al.* (1998) Accurate estimation of travel times from single-loop detectors. *Transportation Research: Part A*, v. 32, n. 1, p. 1–17.
- PTV (2010) *VISSIM 5.30 User Manual*. Planung Transport Verkehr AG – PTV Vision, Karlsruhe, Germany, November 2010.
- Schultz, G. G. (2003) Developing a methodology to account for commercial motor vehicles using microscopic traffic simulation models. 332 p. Tese (Doutorado) – Texas A&M University, College Station, Texas, EUA.
- Zeeshan, R. A. e B. R. Hellinga (2008) Use of Microsimulation to Model Day-to-Day Variability of Intersection Performance. *Transportation Research Record* v. 2088, p.18–25.