FÁBIO GLAUCO WERMERSCH

UMA NOVA ABORDAGEM ENERGÉTICA DA MOBILIDADE URBANA

Tese apresentada ao Programa de Pós-Graduação em Energia do Instituto de Energia e Ambiente da Universidade de São Paulo para a obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. José Goldemberg

Versão Corrigida

SÃO PAULO

2018

Nome: Fábio Glauco Wermersch

Título: Uma Nova Abordagem Energética da Mobilidade Urbana

Tese apresentada ao Programa de Pós-Graduação em Energia do Instituto de Energia e Ambiente da Universidade de São Paulo para a obtenção do título de Doutor em Ciências.

Aprovado em:

Banca Examinadora

Prof. Dr.....Instituição:.....

Julgamento:.....Assinatura:.....

Prof. Dr.....Instituição:.....

Julgamento:.....Assinatura:.....

Prof. Dr.....Instituição:.....

Julgamento:.....Assinatura:.....

Prof. Dr.....Instituição:.....

Julgamento:.....Assinatura:.....

Prof. Dr.....Instituição:.....

Julgamento:.....Assinatura:.....

Prof. Dr.....Instituição:.....

Presidente:.....Assinatura:.....

AGRADECIMENTOS

À minha família, pelo apoio e incentivo.

Ao meu orientador José Goldemberg, sempre atencioso e disponível.

Aos meus supervisores estrangeiros Michael Batty e Elsa Arcaute, por terem me aceitado à realização do estágio sanduíche.

Ao Anastasio Noulas, por ter fornecido os dados das redes sociais para uso neste estudo.

Ao Governo Britânico, pela concessão de uso dos dados de pesquisas origem-destino de Londres.

À UITP-*International Association of Public Transportation*, Bruxelas, pelo fornecimento da publicação *Millenium Cities Database* 2001.

Às colegas Adriana, Júlia e Juliana, da secretaria de pós-graduação e à colega Jussara, sem as quais o estágio sanduíche não seria possível.

Aos amigos Ana Paula, Elena, Flávia, Irene, Joana, Vilma, Dilvan, Duncan e João, pelo apoio.

À Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelas bolsas de estudo, doutorado e sanduíche, concedidas.

RESUMO

WERMERSCH, Fábio G. Uma Nova Abordagem Energética da Mobilidade Urbana. 2018, Tese de Doutorado – Programa de Pós-Graduação em Energia da Universidade de São Paulo, São Paulo, SP, 2018.

A motivação desse trabalho se dá pelo impacto negativo ao meio ambiente dos sistemas de transporte urbano, devido ao consumo de energia. Com o objetivo de encontrar diretrizes sustentáveis para a melhor prática de mobilidade urbana, em termos de consumo de energia, foi desenvolvida uma nova abordagem, utilizando dados desagregados de viagens intraurbanas em 8 regiões metropolitanas mundiais, provenientes de redes sociais digitais. Para a atribuição do modo de transporte utilizado nas viagens nessas oito diferentes localidades, coletadas das redes sociais, foram empregados algoritmos de Aprendizado de Máquina ajustados a dados mesclados de duas pesquisas origem-destino, a de Londres e a de São Paulo. A cidade de menor consumo de energia devido à mobilidade urbana foi Nova Iorque e a de maior consumo foi Los Angeles. Da análise espacial conduzida constatou-se a existência de relação entre a densidade de locais da cidade e a energia média consumida em viagens para esses locais. Também se constatou a predominância da característica urbana de monocentrismo associado a abrangente rede de transporte rápido de massa como diretrizes sustentáveis da melhor prática observada, a cidade de Nova Iorque. Entretanto, apenas o monocentrismo tomado isoladamente pode não ter relação a um menor consumo energético devido à mobilidade urbana, uma vez que se observou o policentrismo na terceira cidade de menor consumo energético, Rio de Janeiro, dentre as oito enfocadas. Para fins de planejamento urbano, o ferramental desenvolvido neste estudo, com uso de dados provenientes de redes sociais, demonstra - de modo mais ágil e barato que as convencionais pesquisas o-d - a possibilidade de obtenção de bases de dados similares às pesquisas origem-destinos convencionais em localidades distantes onde não existem tais pesquisas o-d ou tais pesquisas não sejam fornecidas. Ainda, a observação sobre como se dá a distribuição espacial do consumo de energia devido à mobilidade urbana sobre a área de uma cidade, distribuição com fina granularidade espacial e vinculada a reais deslocamentos observados no espaço urbano - como a que foi propiciada com o ferramental desenvolvido neste estudo -, abre possibilidade a diversas análises de impactos e melhorias de implantação e/ou alteração de facilidades de transportes no ambiente construído.

ABSTRACT

WERMERSCH, Fabio G. A New Energy Approach to Urban Mobility. 2018, Doctoral Thesis - Postgraduate Program in Energy, University of São Paulo, São Paulo, SP, 2018

The motivation of this work is the negative impact to the environment of the urban transport systems due to the consumption of energy. With the objective of finding sustainable guidelines for the best urban mobility practice in terms of energy consumption, a new approach was developed, using disaggregated intra-urban travel data in 8 metropolitan regions worldwide, from digital social networks. For the attribution of the transport mode used in the trips in these eight different locations, collected from the social networks, Machine Learning algorithms were adjusted to the data merged from two origin-destination surveys, London and São Paulo. The city with the lowest energy consumption due to urban mobility was New York and the one with the highest consumption was Los Angeles. From the spatial analysis conducted, it was verified the existence of a relation between the density of localities of the city and the average energy consumed in trips to these places. It was also verified the predominance of the urban characteristic of monocentrism associated with the comprehensive network of fast mass transport as sustainable directions of the best observed practice, the city of New York. However, only isolated monocentrism may not be related to lower energy consumption due to urban mobility, since polycentrism was observed in the third city of lower energy consumption, Rio de Janeiro, among the eight focused. For urban planning purposes, the tool developed in this study, using data from social networks, demonstrates - in a more agile and inexpensive way than conventional surveys o-d - the possibility of obtaining databases similar to conventional origin-destination surveys in distant locations where no such surveys exist, or such surveys are not provided. Also, the observation about the spatial distribution of energy consumption due to urban mobility over the area of a city, distribution with fine spatial granularity and linked to actual displacements observed in urban space - such as that provided with the developed tooling in this study - opens the possibility to several analyses of impacts and improvements of implantation and / or alteration of transport facilities in the built environment.

SUMÁRIO

1 INTRODUÇÃO.....	1
1.1 OBJETIVOS	3
1.2 JUSTIFICATIVA DA ORIGINALIDADE DO TRABALHO	4
1.3 HIPÓTESES A SEREM ESTUDADAS	5
2 MOBILIDADE URBANA E ENERGIA.....	8
3 MÉTODO.....	11
4 DADOS - COLETA E ANÁLISE	14
4.1 DADOS DE CHEGADAS (<i>CHECK-INS</i>) EM DIVERSAS CIDADES MUNDIAIS ..	14
4.1.1 Filtragem dos dados	19
4.1.2 Atributos de entrada	21
4.1.3 Tamanho das amostras de viagens das cidades.....	27
4.1.4 Escolha das cidades a serem enfocadas neste estudo	27
4.2 PESQUISAS ORIGEM-DESTINO	29
4.2.1 Definição das Alternativas de Modos de Transporte	31
5 OBTENÇÃO DO CONSUMO DE ENERGIA DA MOBILIDADE URBANA NAS CIDADES – APRENDIZADO DE MÁQUINA.....	34
5.1 INTRODUÇÃO À MODELAGEM DE ESCOLHA DO MODO DE VIAGEM.....	34
5.2 APRENDIZADO DE MÁQUINA	35
5.2.1 Árvores de Decisão	37
5.2.2 Redes Neurais Artificiais	38
5.2.3 Máquinas de Vetor de Suporte – Support Vector Machines	39
5.3 EXPERIMENTOS DE APRENDIZADO DE MÁQUINA.....	40
5.3.1 Resultados dos Experimentos de AM	41
5.3.2 Investigação sobre a transferência espacial do modelo preditor de escolha do modo de viagem	45
5.4 CÁLCULO DO ÍNDICE DE CONSUMO DE ENERGIA	52
5.4.1 Resultados – Consumo Médio de Energia Devido à Mobilidade Urbana Observada em cada Cidade	54
6 ANÁLISE ESPACIAL.....	70
6.1 Considerações Iniciais	70
6.2 Análise Espacial.....	72
6.2.1 Análise espacial exploratória	72

6.2.2 Análise de agrupamentos	83
6.2.3 Regressão Espacial	101
7 CONCLUSÃO	114
7.1 Resultados e discussão	114
7.2 Recomendações	118
REFERÊNCIAS	119
ANEXO A – Método ANSELIN(2005)	127
APÊNDICE A - Feições do sistema de transporte e os pontos origem-destino da amostra de viagens provenientes das redes sociais nas 8 cidades enfocadas	128
APÊNDICE B – Estatística descritiva dos atributos: energia média consumida nos destinos e quantidade de locais, ambos inseridos na malha de célula 200 metros	136
APÊNDICE C – Histogramas de distâncias de viagens por modos motorizados de São Paulo e Londres: pesquisas origem-destino e viagens das redes sociais.....	138

LISTA DE FIGURAS

Figura 3.1: Estrutura do método proposto	11
Figura 4.1: Distribuição das cidades fornecidas ao redor do globo	17
Figura 4.2 Exemplo, vista dos tipos de locais de parte da cidade de Yokohama contida na base de dados das redes sociais	23
Figura 4.3 Distribuição mundial das 8 cidades enfocadas neste estudo	29
Figura 4.4 Divisão modal da pesquisa OD2007 de São Paulo capital, para 5 modos de viagem	32
Figura 5.1 Relatório de resultados do 1º experimento de diferentes algoritmos de AM	41
Figura 5.2 Relatório de resultado do 1º experimento de SVM	41
Figura 5.3 Relatório de resultado do 2º experimento de SVM	42
Figura 5.4 Gráfico dos resultados das Redes Neurais Artificiais MLP	43
Figura 5.5 Gráfico dos resultados dos algoritmos Random Forest	44
Figura 5.6 Relatório de resultado de performance dos algoritmos RF	44
Figura 5.7 Relatório de resultado de performance do modelo classificador RF com 500 árvores e quatro classes	45
Figura 5.8 Amostra com iguais quantidade de viagens de São Paulo e Londres desbalanceadas conforme a divisão modal da pesquisa origem destino de São Paulo	47
Figura 5.9 Amostra com iguais quantidades de viagens de São Paulo e Londres desbalanceadas conforme a divisão modal da pesquisa origem destino de Londres	48
Figura 5.10 Gráfico dos resultados dos algoritmos Random Fores treinados com 70% dos dados disponíveis e com desbalanceamento de classes conforme a divisão modal de Londres e São Paulo	49
Figura 5.11 Teste de 30% das instâncias contra RF treinado com 70% das instâncias ambos os conjuntos, teste e treinamento, com classes balanceadas	50
Figura 5.12 Teste de 30% das instâncias contra RF treinado com 70% das instâncias, ambos os conjuntos, teste e treinamento, conforme divisão modal de Londres	50
Figura 5.13 Teste de 30% das instâncias contra RF treinado com 70% das instâncias, ambos os conjuntos, teste e treinamento, conforme divisão modal de São Paulo.....	50
Figura 5.14 Indicadores de consumo de energia devido à mobilidade urbana observada em cada cidade	55
Figura 5.15 Quantidade de viagens da amostra de cada cidade após filtragem para dados nulos e faltantes e quantidade de viagens por modos motorizados utilizadas no cálculo dos consumos médios de energia devido à mobilidade urbana observada em cada cidade.....	55
Figura 5.16 Indicadores de energia devido à mobilidade urbana por todos os modos de viagem, incluído o modo caminhada, observada em cada cidade	56
Figura 5.17 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Nova Iorque	57
Figura 5.18 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Nova Iorque	57
Figura 5.19 Divisão modal estimada pelo modelo de escolha algoritmo RF para Nova Iorque	58
Figura 5.20 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Paris	59
Figura 5.21 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Paris.....	59
Figura 5.22 Divisão modal estimada pelo modelo de escolha algoritmo RF para Paris.....	60
Figura 5.23 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Rio de Janeiro	60
Figura 5.24 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Rio de Janeiro.....	61
Figura 5.25 Divisão modal estimada pelo modelo de escolha algoritmo RF para o Rio de Janeiro	61
Figura 5.26 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Toronto	62

Figura 5.27 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Toronto	62
Figura 5.28 Divisão modal estimada pelo modelo de escolha algoritmo RF para Toronto	63
Figura 5.29 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Jacarta .	63
Figura 5.30 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Jacarta.....	64
Figura 5.31 Divisão modal estimada pelo modelo de escolha algoritmo RF para Jacarta.....	64
Figura 5.32 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Londres	65
Figura 5.33 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Londres.....	65
Figura 5.34 Divisão modal estimada pelo modelo de escolha algoritmo RF para Londres.....	66
Figura 5.35 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de São Paulo	66
Figura 5.36 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de São Paulo.....	67
Figura 5.37 Divisão modal estimada pelo modelo de escolha algoritmo RF para São Paulo	67
Figura 5.38 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Los Angeles	68
Figura 5.39 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Los Angeles.....	68
Figura 5.40 Divisão modal estimada pelo modelo de escolha algoritmo RF para Los Angeles.....	69
Figura 6.1 Vista ampliada dos bairros de Nova Iorque com seu sistema de transportes	75
Figura 6.2 Vista ampliada dos bairros de Nova Iorque com locais de destino de viagens	76
Figura 6.3 Vista em detalhe da malha 200x200 metros sobre sistema de transportes de Nova Iorque	77
Figura 6.4 Vista em detalhe da malha 400x400 metros sobre sistema de transportes de Nova Iorque	77
Figura 6.5 Vista em detalhe da malha 600x600 metros sobre sistema de transportes de Nova Iorque	78
Figura 6.6 Gráfico dos valores da estatística Moran's I calculados para a malha de células quadradas, de medidas 200 por 200 metros, e para a média de energia consumida nas viagens inseridas em cada célula	79
Figura 6.7 Gráfico dos valores da estatística Moran's I calculados por malha para a média de energia consumida nas viagens inseridas em cada célula (significância pseudo valor-p =0.001).....	79
Figura 6.8 Teste de significância da estatística Moran's I para o atribuo energia media consumida por célula da malha de Nova Iorque	80
Figura 6.9 Mapa de Nova Iorque com o atributo energia consumida nas viagens com destino contido em célula da malha 200 por 200 metros	81
Figura 6.10 Mapa de Los Angeles com o atributo quantidade de locais por célula da malha 200 por 200 metros	82
Figura 6.11 Mapa de agrupamentos LISA LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Nova Iorque	84
Figura 6.12 Mapa de índice de significância dos agrupamentos LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Nova Iorque	85
Figura 6.13 Gráfico do indicador LISA Moran's I para os agrupamentos do atributo energia média consumida em viagens com destino na célula, Nova Iorque	85
Figura 6.14 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Nova Iorque	86
Figura 6.15 Mapa de índice de significância dos agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Nova Iorque	87
Figura 6.16 Gráfico do indicador LISA Moran's I para os agrupamentos do atributo quantidade de locais por célula, Nova Iorque.....	87
Figura 6.17 Mapa de agrupamentos LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Los Angeles	88

Figura 6.18 Mapa de índice de significância dos agrupamentos LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Los Angeles	89
Figura 6.19 Gráfico do indicador LISA Moran's I para os agrupamentos do atributo energia média consumida em viagens com destino na célula, Los Angeles	89
Figura 6.20 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Los Angeles	90
Figura 6.21 Mapa de índice de significância dos agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Los Angeles	91
Figura 6.22 Gráfico do indicador LISA Moran's I para os agrupamentos do atributo quantidade de locais por célula, Los Angeles	91
Figura 6.23 Mapa de agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Nova Iorque	92
Figura 6.24 Mapa de índice de significância dos agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Nova Iorque.....	93
Figura 6.25 Gráfico do indicador LISA Moran's I bivariado para os agrupamentos dos atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Nova Iorque	93
Figura 6.26 Mapa de agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Los Angeles	94
Figura 6.27 Mapa de índice de significância dos agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Los Angeles	95
Figura 6.28 Gráfico do indicador LISA Moran's I bivariado para os agrupamentos dos atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Los Angeles	95
Figura 6.29 Relação entre estrutura espacial e a efetividade de transporte público	97
Figura 6.30 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Paris.....	98
Figura 6.31 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Rio de Janeiro	98
Figura 6.32 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Toronto	99
Figura 6.33 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Jacarta.....	99
Figura 6.34 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Londres....	100
Figura 6.35 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, São Paulo.	100
Figura 6.36 Mapa de agrupamentos LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Jacarta	109
Figura 6.37 Mapa de índice de significância dos agrupamentos LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Jacarta	110
Figura 6.38 Gráfico do indicador LISA Moran's I para os agrupamentos do atributo energia média consumida em viagens com destino na célula, Jacarta	110
Figura 6.39 Mapa de agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Jacarta	111
Figura 6.40 Mapa de índice de significância dos agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Jacarta.....	111
Figura 6.41 Gráfico do indicador LISA Moran's I bivariado para os agrupamentos dos atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Jacarta	112
Figura 6.42 Mapa de agrupamento LISA Moran's I do atributo energia média consumida em viagens com destino nas células com sistema de metrô dos bairros enfocados de Nova Iorque	113

LISTA DE TABELAS

Tabela 4.1: Cidades Fornecidas.....	16
Tabela 4.2 Quantidade de locais Foursquare recoletados, exemplo em algumas cidades	19
Tabela 4.3 Exemplos de redução na base de dados das redes sociais por filtragem	20
Tabela 4.4 Mapeamento entre as categorias dos atributos motivos de viagem das pesquisas origen-destino e um padrão a ser utilizado neste trabalho.....	24
Tabela 4.5 Trecho do mapeamento de correspondência entre as categorias Foursquare e o motivo padrão	25
Tabela 4.6 Cidades escolhidas a princípio.....	28
Tabela 5.1 Resultados da aplicação das partições de testes, contendo 30% das instâncias, aos algoritmos RF treinados com as respectivas partições de treinamento, contendo 70% das instâncias.....	49
Tabela 5.2 Percentuais para o desbalanceamento da divisão modal.....	51
Tabela 5.3 Resultados de performance dos algoritmos RF preditores do modo de viagem por cidade.....	52
Tabela 5.4 Exemplo: Consumo de energia por modo de viagem de Paris.....	53
Tabela 5.5 Consumos médios unitários de energia para os modos de viagem a serem utilizados no cálculo do indicador de consumo de viagem devido à mobilidade urbana	54
Tabela 6.1 Sumário de estatística descritiva* dos atributos energia consumida nas viagens com destino na célula e quantidade de locais por célula	83
Tabela 6.2 Estatística Moran's I calculadas para o atributo energia média consumida em viagens com destinos inseridos dentro da célula de 200x200 metros, para as demais seis cidades enfocadas	103
Tabela 6.3 Resultados da regressão, cidade de Nova Iorque	104
Tabela 6.4 Resultados da regressão, cidade de Paris	104
Tabela 6.5 Resultados da regressão, cidade do Rio de Janeiro.....	105
Tabela 6.6 Resultados da regressão, cidade de Toronto	105
Tabela 6.7 Resultados da regressão, cidade de Jacarta.....	106
Tabela 6.8 Resultados da regressão, cidade de Londres	106
Tabela 6.9 Resultados da regressão, cidade de São Paulo.....	107
Tabela 6.10 Resultados da regressão, cidade de Los Angeles	107
Tabela 6.11 Modelos clássicos OLS e espaciais obtidos para a relação dos atributos, conforme as tabelas anteriores, e após a retirada de outliers espaciais, considerando as relações dos atributos que apresentaram o maior coeficiente de determinação	108

LISTA DE ABREVIATURAS E SIGLAS

GPS Global Positioning System

UITP International Association of Public Transport

API Application Program Interface

o-d Origem-Destino

UTM Projeção Universal Transversa de Mercator

OSNG Sistema de Referência Nacional Militar Britânico

WGS84 Sistema Geodésico Mundial revisão 1984

OD2007 Pesquisa origem-destino de São Paulo capital

AM Aprendizado de Máquina

IA Inteligência Artificial

RNA Redes Neurais Artificiais

AD Árvores de Decisão

RF Random Forest

MLP Multilayer Perceptron

SVM Support Vector Machines

ESDA Análise Exploratória de Dados Espaciais

MAUP Problema da Unidade de Área Modificável

LISA Moran's I Local

MRT Transporte Rápido de Massa

OLS modelo Clássico Linear

SAR modelo Simultâneo Autoregressivo

SLM modelo Lag Espacial

BRT Transporte Rápido de Ônibus

MRT Transportes Rápidos de Massa

SEM modelo Erro Espacial

1 INTRODUÇÃO

A partir da revolução industrial, as cidades passaram a ter um sistema de transporte urbano proporcionado por uma transição de combustível e intensificação do uso energia na época, criando, assim, mais um sistema urbano de uso de energia além dos então existentes, sistema de aquecimento e de luz elétrica RUTTER & KEIRSTEAD (2012). Atualmente, segundo a IEA (2015), a maior fatia de uso final de energia global é devida ao setor de transporte, apresentando também a maior estimativa de taxa de crescimento no uso de energia dentre os demais setores da economia, até 2030. As viagens intraurbanas atualmente constituem mais que sessenta por cento de todos os quilômetros viajados globalmente e deverão ser triplicadas até 2050 (VAN AUDENHOVE et al., 2014). Segundo LEFEVRE (2009), o consumo de energia pelo sistema de transporte urbano representa uma grande parcela do consumo total urbano, geralmente entre 20% a 50%, excluído o consumo industrial. Com tal expressão do setor, são observados numerosos impactos negativos que ocorrem em maior intensidade na área urbana - devido à sua inerente característica concentradora -, dentre eles congestionamentos, acidentes, poluição do ar e consumo de energia, o que atrai a atenção para estudos de medidas mitigadoras.

É prevista uma escalada global de crescimento de áreas urbanas sem precedente, acentuada nas regiões em desenvolvimento, com predominância para o continente asiático (CHEN et al., 2014). Alguns estimam que globalmente a quantidade de área urbana pode triplicar entre 2000 e 2030 (SETO et al., 2012).

Um estudo de FOUQUET (2012) modelou através de técnicas econométricas um longo período da demanda por transporte de passageiro, desde 1850 até 2010, da Inglaterra, para estimar suas elasticidades de receita e de preço ao longo do tempo. Dentre os resultados apresentados, ele concluiu que o uso de energia em transportes é guiado pela demanda por mobilidade e que transições de energia e tecnologia não irão parar essa demanda, bem como sua tendência crescente.

Tal asserto de FOUQUET encontra-se em acordo com GEA (2012) que também concluiu sobre a necessidade de políticas estruturais e de demanda, em complemento à inovação tecnológica para eficiência energética, veicular e desenvolvimento de combustíveis alternativos, na direção de um sistema de transporte urbano mais sustentável que atenda aos adicionais três bilhões de moradores urbanos estimados para até 2050.

Em termos de impactos ao meio ambiente, talvez não haja tantos planetas Terra quanto os necessários para que as regiões em desenvolvimento se urbanizem percorrendo toda a curva de aprendizado de urbanização já percorrida pelas regiões ou sociedades desenvolvidas. Seria conveniente ao planeta que houvesse um *leapfrogging* (GOLDEMBERG, 2012) tecnológico, de forma que as melhores práticas fossem identificadas e aplicadas a tempo nessa escalada de urbanização. Nessa direção, a identificação da solução urbana em transportes de pessoas, a mobilidade urbana, que menos demandasse ao planeta em consumo de energia poderia contribuir nesse processo evolutivo global.

Com efeito, o presente momento, de acordo com a literatura consultada, indica, não por acaso, haver uma busca pela identificação tanto da melhor prática em consumo energético por mobilidade urbana, como também a busca pelas causas dessa relação do sistema de transporte e o espaço urbano no qual se insere. Pois, frequentemente, o consumo do sistema de transporte urbano é relacionado com a forma urbana e a interação humana nela, sendo essas consideradas mutuamente relacionadas. A literatura consultada, adiante apresentada, abarca alguns estudos dessas relações, porém o tema ainda é apresentado como uma questão aberta até o presente.

Tais investigações a respeito da mobilidade urbana envolvem como um único elemento amostral uma cidade inteira, qual seja, sua mobilidade urbana coletada e, com efeito, diversas cidades seriam necessárias a compor um espaço de busca por uma eventual melhor prática observada. Porém, um dos aspectos dificultadores a esse intuito é a obtenção da mobilidade urbana coletada em diversas cidades, por exemplo em escala planetária, e, preferencialmente, ocorrida de forma simultânea, de maneira a propiciar uma análise entre as cidades em mesma época tecnológica, seus sistemas e modos de transporte utilizados e os consumos de energia decorrentes.

A forma clássica de coleta da mobilidade urbana de uma cidade, as pesquisas origem-destino, possuem custo elevado de coleta e com raríssimas exceções estão disponíveis para acesso de forma pública e gratuita, tampouco são disponibilizadas mesmo que formalmente

solicitadas para fins de pesquisa por pesquisadores distantes de suas localidades ou países. Para exemplificar a situação, apenas duas cidades foram obtidas de forma gratuita e com a precisão de localização desejada para fins desta pesquisa, São Paulo e Londres.

Felizmente, o surgimento recente, 2009, de redes sociais baseadas em locais, ou seja, uma plataforma de rede social que utiliza a localização como sua principal característica, como por exemplo a rede social *Foursquare*, veio a compor uma nova fonte de dados de mobilidade. Neste tipo de rede social uma entidade virtual é criada, com atribuições de características, e associada, com uma localização geográfica precisa, a um lugar fisicamente existente no mundo real por seus usuários, participantes de forma pública e gratuita - portanto, de maneira conhecida como “*crowdsourcing*”. Tais lugares passam, então, a ser relatados de forma pública ou privada por um usuário toda vez que ele chega em um desses lugares, de forma a que outros usuários saibam sobre sua localização naquele momento. A disponibilidade de coleta de todas as chegadas de todos os usuários de uma cidade por tempo prolongado, meses, proporciona tal nova fonte de dados de mobilidade, ainda que com viés estatístico a diante neste texto discutido.

Com o propósito de apresentar nova contribuição à literatura, este trabalho apresenta a observação simultânea da mobilidade urbana em diferentes metrópoles mundiais, através da coleta de viagens intraurbanas realizadas por indivíduos e por eles informadas por meio de seus aparelhos telefônicos móveis, para a obtenção de estimativa do consumo de energia por transporte. A partir de tal observação, estima-se qual cidade apresenta o menor consumo energético relativo à mobilidade urbana observada, ou em outras palavras, qual cidade apresenta a mobilidade urbana mais sustentável sob o ponto de vista energético ao planeta, à procura de diretrizes sustentáveis.

Sendo assim, espera-se que a realização desse trabalho seja um expoente da possibilidade de observação simultânea e remota da mobilidade urbana, neste caso na dimensão energética, em escala global com uso da tecnologia do presente, bem como que, a partir da melhor prática encontrada, se possa contribuir para a estipulação de diretrizes sustentáveis a serem exploradas ou reproduzidas na crescente escalada de urbanização atual.

1.1 OBJETIVOS

Obtenção de um ordenamento das metrópoles sob estudo em função do consumo energético calculado a partir da mobilidade urbana observada com, portanto, a indicação da

metrópole de menor consumo energético, ou melhor eficiência, em sua mobilidade. Tal ordenamento foi obtido a através da elaboração de instrumento único de cálculo, planilha, do consumo energético de mobilidade urbana para cada espaço urbano, metrópole, a partir de dados tabulados de forma padronizada. Para tanto, procurou-se estimar a divisão modal a cada localidade/metrópole estudada, de forma a aproximar os deslocamentos coletados para uma quantificação comparativa do consumo de energia entre as cidades.

Buscou-se, também, propiciar uma indicação à literatura pertinente a respeito da metrópole com menor estimativa do consumo energético de mobilidade urbana, e suas características ou diretrizes sustentáveis. Desta forma, espera-se contribuir em área de políticas públicas para a estipulação de quais são as razões subjacentes ao processamento da respectiva mobilidade urbana com menor consumo energético e suas diretrizes sustentáveis observadas, com vistas à uma possível reprodução dessas diretrizes sustentáveis na crescente urbanização mundial, para um menor impacto/demanda ao globo terrestre, devido à mobilidade urbana.

1.2 JUSTIFICATIVA DA ORIGINALIDADE DO TRABALHO

A nova abordagem ora proposta é a obtenção de um ranque em consumo de energia devido à mobilidade urbana de cidades em perspectiva mundial, através da mobilidade voluntariamente publicada em domínio público digital, *Internet*, por seus cidadãos, contemplando simultaneidade na coleta da mobilidade em diversas cidades globalmente distribuídas.

Há na literatura diversas menções sobre a custosa e morosa forma convencional de se obter dados sobre a mobilidade urbana de uma cidade, geralmente realizada com periodicidade maior que anual, e sobre o surgimento de novas formas de obtenção, como por exemplo, através do rastreamento de telefones móveis equipados com GPS (AXHAUSEN, 1998; ETTEMA et al., 1996; JARIYASUNANT et al., 2011; BARTHELEMY, 2011; STENNETH et al., 2011; TEEUW et al. 2011; COTTRILL et al., 2013; FENG & TIMMERMANS, 2013; WANG et al., 2013; IQBAL et al., 2014).

Para fazer frente à demanda ambiental ao globo terrestre impelida pela iminente escalada de urbanização sem precedente e já iniciada, será necessário nas regiões em urbanização adotar as melhores soluções em sistemas urbanos existentes a época, de forma a haver um *leapfrogging* tecnológico para o benefício ambiental terrestre. Ocorre que, no

presente, conforme literatura pesquisada, há uma busca pela descoberta de melhores práticas e soluções e, salvo engano deste autor, há poucos estudos com abrangência mundial disponíveis sobre o consumo de energia em diferentes cidades mundiais, KENWORTHY & LAUBE (2001) sobre dados desagregados e LE NÉCHET (2012) sobre dados agregados.

Porém, o estudo feito por KENWORTHY & LAUBE (2001) apresenta a morosidade para a obtenção e atualização de dados desagregados das diferentes cidades mundiais, que, a propósito, na publicação KENWORTHY (2014) apresentava atualização de dados datada em 2005 e 2006, ou seja, defasada a quase uma década.

É evidente a necessidade de outros trabalhos na mesma direção e que possuam a agilidade esperada para a atual necessidade humana. Nesse contexto se aplica o trabalho proposto pela busca inédita de alternativa à convencional obtenção de informações de viagens intraurbanas como as realizadas anteriormente, para fins de análise comparativa de abrangência mundial sobre o que há em termos de consumo de energia da mobilidade urbana.

Assim, justifica-se a relevância da proposta deste trabalho por sua nova abordagem quantitativa e desagregada, que emprega a coleta de dados abrangente nas dimensões espaço, tempo e universo de viagens efetivamente realizadas e coletadas simultaneamente ao redor do globo, com o emprego de tecnologia recentemente criada. Tais características estão frequentemente mencionadas como desejosas a trabalhos futuros na literatura consultada e citada na introdução deste texto. Também, as diretrizes sustentáveis urbanas almejadas neste estudo, que visam a otimização das estruturas urbanas de forma a reduzir o impacto global pelas crescentes cidades mundiais, na esperança da obtenção de um *leapfrogging* nessa atual franca expansão urbana, são frequentemente mencionadas na literatura (GEA, 2012; EVI, 2016).

1.3 HIPÓTESES A SEREM ESTUDADAS

O trabalho realizado se apoia na premissa de que há uma norma universal de mobilidade urbana apresentada no recente estudo de NOULAS et al. (2012), o qual coletou dados da mobilidade urbana em diferentes metrópoles mundiais e obteve o ajuste de uma única função ao comportamento de viagem para todas as 34 metrópoles observadas. Os autores apresentam que as variações nos movimentos humanos nas diferentes cidades são predominantemente devidas às diferentes distribuições dos locais onde nossas atividades cotidianas ocorrem, entre diferentes ambientes urbanos. Estudo recente, WILKERSON et al. (2014), abrangendo 76

cidades da Alemanha e utilizando dados de pesquisas convencionais sobre viagens e, portanto, mais completo quanto a quantidade de atributos disponíveis, citando o trabalho de NOULAS et al. (2012), levanta questão sobre a possibilidade da consideração de diferentes modos de viagem, ou hora do dia, ou dia da semana, poder gerar diferentes padrões universais.

Usando, então, uma base de dados bem mais completa quanto à diversidade de informações sobre a viagem e o viajante, e eventualmente com viés mais homogêneo por se tratar de cidades constituintes de um mesmo país, os autores do estudo de WILKERSON et al. (2014) apresentaram que seu melhor ajuste obtido se aproximou muito daquele obtido anteriormente por NOULAS et al. (2012), e também que a resposta indeterminada do comprimento de viagem em relação à população das cidades que observaram pode dar suporte à independência entre o comprimento de viagem e área da cidade, conforme NOULAS et al. (2012) havia observado.

Ainda a respeito dessa primeira premissa de que há uma norma universal de mobilidade urbana, BARTHELEMY (2011) apresenta que estudos anteriores - como KOELBL & HELBING (2003) o qual estabelece que a energia consumida por indivíduos para viajar é constante e independe do modo de viagem, e, por sua vez, HUBERT & TOINT (2007) mostram que a distribuição do tempo de viagem normalizado parece ser a mesma para três diferentes países estudados - já apontavam para a possível existência de características universais do movimento humano.

Foge ao escopo do presente estudo uma discussão sobre essa norma universal de mobilidade a qual foi tema central dos estudos anteriores, acima mencionados, os quais a dão suporte, sendo que neste estudo tal norma é tomada de início como primeira premissa. Mesmo assim, destaca-se o fato de que o estudo recente de NOULAS et al. (2012) descreve confirmar em escala mundial a teoria de STOUFFER (1940), o que dá indício de alguma estabilidade dessa norma comportamental humana devido ao tempo de mais de meio século decorrido de sua proposição até tal confirmação com tecnologia recente.

A segunda premissa de apoio deste trabalho realizado é a possibilidade de cálculo da estimativa do consumo de energia dos deslocamentos intraurbanos coletados a partir da *Internet*, em diferentes cidades mundiais por uma única formulação, ou único ferramental.

De acordo com o realizado neste trabalho, o objetivo do cálculo de consumo de energia devido à mobilidade de uma cidade não é a quantificação exata da energia consumida de cada

uma das cidades coletadas, mas sim, a obtenção de um fator comum que permitiu uma comparação entre elas, apontando qual delas apresentou a melhor eficiência para a mobilidade observada, ou a mobilidade mais sustentável.

O foco é dado na importância do consumo de energia como um adequado indicador geral de sustentabilidade, como apontado no relatório de Brundtland (WCED, 1987), uma vez que todos os tipos de energia, renováveis ou não, causam um grande espectro de impactos e, portanto, uma ameaça à sustentabilidade ao longo prazo (HOLDEN et al., 2013).

Sendo assim, conclui-se haver a possibilidade de obtenção da estimativa do consumo de energia em transporte para a mobilidade urbana observada em cada diferente metrópole.

Paralelamente à proposição principal deste estudo formalizada nas premissas e conclusão acima descritas, uma hipótese inerente ao desenvolvimento do estudo foi a do uso das redes sociais para obtenção de pesquisas origem-destino similares.

Devido à dificuldade de obtenção de pesquisas origem-destino em escala mundial, é proposto neste estudo o uso de redes sociais para um mesmo fim ao qual as pesquisas origem-destino são utilizadas, quer seja observação da mobilidade urbana em determinada cidade. Ainda que os dados provenientes das redes sociais empregados neste estudo não contenham a mesma riqueza de informações de uma pesquisa o-d, pode-se revelar útil tal utilização das redes sociais como um substituto similar em diversos estudos que venham a empregar pesquisas o-d em localidades distantes onde não se consegue obter tais pesquisas o-d, ou tais não sejam fornecidas, ou ainda, onde não existem pesquisas o-d.

2 MOBILIDADE URBANA E ENERGIA

Este capítulo apresenta breves descrições de trabalhos anteriores realizados na mesma linha de pesquisa a qual se insere o presente estudo: a relação do ambiente construído e o consumo de energia devido à mobilidade urbana. Nesse sentido, foram selecionados estritamente trabalhos anteriores que mais se aproximam do estudo ora proposto, dentre os encontrados na literatura pertinente, bem como feita uma breve crítica geral a respeito do ponto nessa linha de pesquisa ao qual tais trabalhos conduzem.

O estudo conduzido por NEWMAN & KENWORTHY (1989a, 1989b) explorou a relação entre densidade urbana e uso de energia em transporte. O resultado apresentou uma forte correlação negativa entre densidade populacional e o consumo de energia em transporte per capita, a partir de um conjunto de dados de cidades em diferentes países. Entretanto, MIDALI et al. (2004) usou o mesmo conjunto de dados e concluiu que não havia impacto direto da densidade urbana nas distâncias viajadas. Ainda, GOMEZ-IBANEZ (1991) aponta a falta de controle estatístico das variáveis socioeconômicas utilizadas como uma falha do resultado apresentado por NEWMAN & KENWORTHY (1989a, 1989b).

Outro estudo foi realizado por BANISTER et al. (1997) em que são abordados casos sobre cinco cidades britânicas e uma holandesa, no qual os resultados mostram relações significativas entre o consumo de energia em transportes e as características físicas da cidade, tais como densidade, tamanho e quantidade de espaços vazios.

A partir de dados detalhados de viagem (*Dutch National Travel Survey*) para estimar o consumo de energia em transporte somente por motivo trabalho, através de modelos de regressão, na Holanda, SUSILO & STEAD (2008) apresentam resultados indicadores de que

viajantes pendulares residentes em áreas urbanas mais densas consomem menos energia do que aqueles que moram em áreas menos urbanizadas.

BROWNSTONE & GOLOB (2009) modelaram a relação entre densidade residencial, uso de veículo e consumo de combustível para domicílios da Califórnia. Embora tenham empregado somente densidade residencial para descrever o ambiente construído, eles concluíram que um aumento de 40% na densidade residencial está associado a uma redução de 5% na distância viajada por automóvel a partir do domicílio.

PARSHALL et al. (2010) modelaram o consumo de energia e emissões para área urbana e rural nos EUA a partir de um inventário nacional de consumo de energia chamado Vulcan, sendo a participação devida a transportes calculada apenas a partir de informação agregada do consumo de combustíveis fósseis inventariado. Uma limitação do trabalho foi a impossibilidade de calcular o consumo de energia ou emissões para áreas urbanas individuais, cidades, devido à dificuldade de alocar os dados de células do inventário que interceptam mais de uma área urbana.

LIU & SHEN (2011) também utilizaram modelos matemáticos para abordar a relação entre a forma urbana e o consumo de energia em transporte gerado pelos veículos particulares de domicílios de uma região metropolitana norte-americana. Os resultados apontam que, dentre as características físicas da cidade tomadas, a acessibilidade provê maior capacidade explicativa do que a densidade, o formato e a diversidade.

ECHENIQUE et al. (2012) abordaram o paradigma da forma de cidade compacta como a mais vantajosa em termos socioeconômicos e ambientais. Utilizando modelos estatísticos para três regiões britânicas, simularam respostas comportamentais frente a hipotéticas políticas dispostas em vinte e seis critérios, socioeconômicos e ambientais, inclusive a demanda de energia por transporte. Concluíram que nenhuma forma urbana foi claramente superior e que a influência da forma urbana no consumo de energia em transporte é muito modesta.

LE NÉCHET (2012) utilizou dados agregados censitários populacionais dispostos em malha de células de 200 metros, cruzados com dados agregados sobre mobilidade contidos em UITP(2001) de 34 cidades mundiais, na tentativa de avaliar o poder explanatório de indicadores socioeconômicos e morfológicos na busca de formas urbanas sustentáveis. Um dos indicadores de mobilidade urbana utilizado foi o consumo de energia por habitante devido a transportes.

Sua conclusão foi que os indicadores socioeconômicos e morfológicos são complementares na busca intentada por ele.

Esses estudos que abordam em algum grau a mobilidade urbana e seu correspondente consumo de energia, encontrados na literatura consultada e nos casos supracitados, ou utilizam dados agregados, ou dados sobre apenas uma parcela dos viajantes intraurbanos - por exemplo, somente os que viajam por motivo trabalho, ou somente por modo automóvel particular - ou faltam atributos sobre as viagens coletadas, ou estimam o consumo de energia por presumida correlação da qualidade do ar na área urbana, ou estimam o consumo de energia por distância viajada apenas por modo automóvel. Ainda, alguns dos fatores descritos acima foram apresentados isoladamente ou em conjunto no mesmo estudo, deixando para trabalhos futuros abordagens de coleta de dados e/ou estimativas ou medidas do consumo de energia que melhor se aproximem da realidade.

O trabalho com semelhante abrangência mundial e com dados desagregados como o ora proposto foi o de KENWORTHY & LAUBE (2001), de acesso indisponível de forma gratuita, o qual tem uma visão geral de seus resultados publicados em KENWORTHY & LAUBE (2002) e em NEWMAN & KENWORTHY (2011), bem como uma atualização de base de dados, originalmente de 1995 para 2005 e 2006, com redução de abrangência para quarenta e uma cidades em KENWORTHY (2014). Em KENWORTHY & LAUBE (2002), a partir de dados reunidos ao longo de três anos sobre oitenta e quatro cidades ao redor do globo, os autores apresentam uma perspectiva básica sobre onde tais cidades estão situadas em aspectos relativos a transportes, uso do solo e economia e ambiente. Esse trabalho calculou o consumo de energia devido à mobilidade urbana das cidades enfocadas, e embora não seja detalhada a metodologia utilizada nas publicações de acesso público gratuito supracitadas, provavelmente utilizou dados provenientes das formas convencionais de coleta de dados para matrizes origem-destino em transportes.

3 MÉTODO

Para o atingimento dos objetivos deste trabalho foi elaborado um método em quatro etapas: filtragem e preparação dos dados de deslocamentos; obtenção da divisão modal por localidade e atribuição do modo de viagem; atribuição do consumo unitário do modo de viagem a cada uma das viagens; e, por fim, o cálculo do consumo de energia por mobilidade urbana em cidade sob estudo e apresentação do ranque das cidades por ordem de consumo. A partir da obtenção do ranque das cidades obtido, realiza-se uma análise espacial exploratória sobre a(s) cidade(s) de menor consumo energético em busca de diretrizes sustentáveis.

Figura 3.1: Estrutura do método proposto

O ordenamento almejado a respeito da eficiência energética das mobilidades urbanas observadas foi obtido através da medida de energia gasta por passageiro quilômetro viajado, J/pkm, conforme cada modo de viagem utilizado. Este trabalho se limitou a obter quantidade de energia final (GOLDEMBERG, 2012) utilizada à locomoção de passageiros, deixando fora do seu escopo a conversão dessa em energia primária, e suas prováveis características locais à conversão.

Após um primeiro processo inerente a utilização de qualquer grande base de dados como filtragem, estatística descritiva e eventual mudança de formato para processamento nas etapas seguintes, obteve-se a divisão modal para cada cidade. Para tal tarefa foram empregadas técnicas computacionais de extração de conhecimento da divisão modal das cidades com as informações disponíveis, pesquisas O-D de São Paulo e Londres, à semelhança do já realizado por este autor em WERMERSCH (2002). Nesse processo de obtenção do modelo de escolha com utilização de algoritmos de aprendizado de máquina, os atributos receberam tratamento conforme seu tipo. Atributo discreto nominal, ou seja, sem ordenação, como por exemplo o atributo motivo de viagem foi transformado em vetor de valores binários, onde somente um elemento do vetor é igual a 1 e os demais são iguais a 0, variável tipo *dummy*. Por sua vez, os atributos numéricos foram normalizados para um mesmo intervalo de 0 a 1, pois dados com atributos em escalas diferentes podem confundir o algoritmo de aprendizado de máquina, fazendo com que ele dê mais importância para atributos com maiores valores (MARTINELLI, 1999). Entretanto, cada atributo numérico deve ser normalizado individualmente, a partir dos próprios valores limites observados entre suas instâncias.

A partir do conhecimento extraído sobre o comportamento de escolha do modo de viagem, modelo ajustado, foi realizada a predição dos modos de transporte para as viagens de outras diferentes cidades, ou seja, realizou-se a translação espacial do modelo ajustado com dados de São Paulo e Londres para outras cidades. De forma a propiciar a translação do modelo de escolha do modo de viagem, apenas os atributos disponíveis na base de dados proveniente das redes sócias, ou seja, apenas atributos de viagem, foram utilizados para o ajuste do modelo de escolha do modo de viagem, embora houvesse mais atributos disponíveis nas pesquisas O-D obtidas. Para a translação do modelo foi empregada em sua fase de ajuste, ou aprendizado, informação sobre a divisão modal da cidade de aplicação do modelo, com vistas a uma melhor performance preditiva do modelo ajustado.

Para o processo de atribuição do consumo energético das viagens em cada cidade, foram obtidas as informações sobre os consumos de energia unitários de cada modo de transporte, característico de cada cidade, em publicação da União Internacional de Transportes Públicos, UITP (2001).

Com o ordenamento das cidades em termos do consumo por mobilidade, já na fase de discussão dos resultados deste trabalho, buscou-se diretrizes sustentáveis contidas naquela(s) cidade(s) de menor consumo energético devido à sua(s) mobilidade(s) urbana, por meio de análise espacial. Uma análise comparativa entre as características da mobilidade urbana observada, relacionada a seu consumo de energia na(s) cidade(s) de menor consumo e aquela de maior consumo, buscou o atendimento de um dos objetivos deste trabalho, o de contribuir para a orientação de políticas públicas sustentáveis e referidas como necessárias conforme COP21 em 2015 (UNFCCC, 2015). Ou seja, advém do caráter exploratório desta pesquisa a busca por razões subjacentes a um processamento da mobilidade urbana mais energeticamente eficiente.

4 DADOS - COLETA E ANÁLISE

Este capítulo contém a descrição pormenorizada dos procedimentos da fase de aquisição do conjunto de dados empregados neste estudo. O conjunto de dados utilizado é composto de diversas origens e formatos. Informações sobre momentos de chegadas em locais de diversas cidades provenientes de redes sociais, coleta de informações sobre a categoria desses locais, se restaurante ou residência, dados de viagens de pesquisas origem-destino de duas grandes capitais mundiais, coleta de mapas das feições urbanas de diversas cidades e coleta de distâncias de viagens através de aplicativo da *Internet*, o *Google Maps*.

4.1 DADOS DE CHEGADAS (*CHECK-INS*) EM DIVERSAS CIDADES MUNDIAIS

Os dados a serem utilizados neste trabalho foram coletados na pesquisa de doutorado NOULAS (2013) e foram fornecidos por ele para fim exclusivo de desenvolvimento desta pesquisa de doutorado por este autor.

Esses dados foram coletados durante os meses de Janeiro a Maio de 2011 da publicação voluntária da localização dos usuários de telefonia móvel da rede social chamada *Foursquare* na rede social de domínio público chamada *Twitter*, vide NOULAS et al. (2012 e 2013). São constituídos por informações de chegada em determinados locais urbanos contendo as coordenadas latitude-longitude precisamente obtidas através do equipamento GPS dos telefones móveis de usuários, ou pelo usuário, que previamente cadastrou aquele local, a hora exata daquele momento de chegada e uma tipificação a que se refere o local se, por exemplo “Loja de Café”, “Casa/Residência”, “Shopping/Mall”, “Supermercado”, “Empresa” ou “Restaurante/Alimentação”, além do nome real do estabelecimento quando aplicável.

Alguns usuários da rede social *Foursquare* fazem a publicação de sua chegada no lugar de forma espontânea na outra rede social de domínio público Twitter, quando então tornam-se de conhecimento público e passíveis da coleta realizada. O autor dessa coleta de *check-ins* descreve em suas publicações (NOULAS et al., 2012, 2013) que a quantidade de dados coletados a partir do total disponível pelo proprietário da rede social *Foursquare* em uma localidade foi entre vinte e vinte e cinco por cento. É interessante observar a demonstração em vídeo disponível na *Internet*, vide links abaixo, do pulso diário de mobilidade de uma cidade gerado pela quantidade de deslocamentos/viagens da qual *Foursquare* dispõe.

<https://pt.foursquare.com/infographics/pulse> (Escolher lá a cidade a visualizar e clicar no canto esquerdo do vídeo para iniciá-lo.)

<https://www.youtube.com/watch?v=wrInToGwiZQ> (Em outro *site*, por exemplo.)

Os *check-ins*, as chegadas em locais da cidade publicadas, coletados podem ou não vir a constituir um deslocamento intraurbano, ou viagem, fulcro deste estudo, a depender das características entre dois *check-ins* que possam torna-los origem e destino de uma viagem.

A forma de coleta de dados empregada da base de dados fornecida traz vantagens e desvantagens em relação à forma convencional de coleta de viagens intraurbanas, as pesquisas origem-destino por entrevista domiciliar. Há precisão elevada, por GPS, na localização dos pontos de *check-in*, eventuais pontos de origem e destino, contrariamente às pesquisas convencionais, nas quais o entrevistado declara por endereços suas viagens realizadas no dia anterior. A quantidade amostrada também supera muito a das pesquisas convencionais.

A periodicidade ou disponibilidade contínua de plataformas, ou redes sociais, da *Internet* para a coleta de informações por qualquer pessoa interessada para a formação de uma base de dados - como a fornecida - traduz-se em agilidade na observância do fenômeno desejado ou monitoramento da mobilidade urbana.

Porém, ao menos por enquanto, em dados provenientes das redes sociais há a falta de informação sobre o modo de viagem utilizado, e é um viés inerente o fato de que apenas uma parcela da população possui os telefones móveis com *GPS* embutido, chamados *smartphones*. Apenas os *smartphones* são capazes de operar o *software* necessário às plataformas sociais geolocalizadas (*Foursquare*) e de publicação de informações em domínio público (*Twitter*). Além disso, nem todos os usuários da rede social *Foursquare* realizam a publicação em rede

social de domínio público, publicação necessária que possibilita a coleta como a realizada. Ainda, é desfavorável a impossibilidade de serem obtidos deslocamentos intermediários entre uma publicação de uma localização chegada, *check-in*, e a seguinte.

Um provável fator está relacionado ao viés desse conjunto de dados proveniente de redes sociais, vinculado a posse de aparelho telefônico móvel mais sofisticado, assim considerado na época da coleta, 2011: a capacidade econômica do viajante intraurbano, fator provavelmente variado com mais ou menos intensidade, dependendo da localização no globo terrestre. Afora este fator de provável estratificação prévia da amostra, a considerar o contexto de quais indivíduos publicam ou não suas chegadas e quando o fazem, é esperada a aleatoriedade da coleta destes eventos e, por conseguinte, das viagens.

Contudo, todas as formas de coletas de dados atuais possuem vieses, e alguns dos vieses mencionados na literatura sobre a coleta de dados através de aparelhos telefônicos móveis tendem a desaparecer com o tempo (REDDY et al., 2010; XU et al., 2011; COTTRILL et al., 2013; FENG & TIMMERMANS, 2013; SAFI et al., 2013; WANG et al., 2013; IQBAL et al., 2014).

Foram fornecidos dados referentes a cem cidades, globalmente espalhadas, vide tabela e figura abaixo:

Tabela 4.1: Cidades Fornecidas

Cidades Fornecidas							
1	Amsterdam	26	Fukuoka	51	Malang	76	San Antonio
2	Anaheim	27	Gapyeong Gun	52	Manado	77	San Diego
3	Atlanta	28	George Town	53	Manila	78	San Francisco
4	Austin	29	Honolulu	54	Medan	79	San Jose
5	Baltimore	30	Houston	55	Melbourne	80	Santiago
6	Bandung	31	Huai-khwang	56	Mexico City	81	São Paulo
7	Bang Na	32	Incheon	57	Miami	82	Seattle
8	Barcelona	33	Indianapolis	58	Milano	83	Selangor
9	Belo Horizonte	34	Istanbul	59	Milwaukee	84	Semarang
10	Bogor	35	Jakarta	60	Minneapolis	85	Sentosa Island
11	Boston	36	Johor Bahru	61	Moscow	86	Seoul
12	Brasilia	37	Jurong-west	62	Nagoya	87	Shah Alam
13	Brussels	38	Kawasaki	63	New York	88	Singapore
14	Buenos Aires	39	Khlong Toei	64	Orlando	89	Surabaya
15	Charlotte	40	Kuala Lumpur	65	Osaka	90	Sydney
16	Chatuchak	41	Kulon Progo	66	Paris	91	Tampa

17	Chiang Mai	42	Kuta	67	Pathum Wan	92	Tampines
18	Chicago	43	Kyoto	68	Petalang Jaya	93	Tangerang
19	Cleveland	44	Las Vegas	69	Philadelphia	94	Tokyo
20	Columbus	45	Lima	70	Phoenix	95	Toronto
21	Curitiba	46	London	71	Pittsburgh	96	Utrecht
22	Dallas	47	Los Angeles	72	Portland	97	Vancouver
23	Denpasar	48	Madrid	73	Porto Alegre	98	Washington
24	Denver	49	Makassar	74	Quezon City	99	Yogyakarta
25	Fortaleza	50	Makati City	75	Rio de Janeiro	100	Yokohama

Fonte: o próprio autor do texto.

Figura 4.1: Distribuição das cidades fornecidas ao redor do globo

Fonte: o próprio autor do texto.

Estes dados fornecidos são separados em duas partes, dois arquivos digitais:

- em uma parte, estão dispostos os dados de forma não tabulada contendo códigos de locais associados a datas com hora que correspondem às chegadas e essas associadas a outro código aleatório que corresponde ao usuário da rede social, correspondendo à forma bruta da coleta realizada em quatro meses;
- e em outra, estão dispostos os códigos de locais associados a algumas de suas características ou atributos, como por exemplo suas coordenadas globais (latitude, longitude), uma categoria conforme categorizado na rede social *Foursquare*, um nome, etc, também de forma não tabulada.

Para lidar com a primeira metade dos dados fornecidos que contém as informações de chegadas, foi elaborado um programa em linguagem computacional que pudesse identificar um usuário e a sequência de chegadas correspondentes a ele, concatenando em pares por ordem cronológica e dispondo tais pares de forma tabulada. Da análise dos dados após serem tabulados notou-se que grande parte deles não figurariam como uma viagem intraurbana, pois embora pertencessem a um mesmo usuário e fossem sequenciados por data e hora, havia em muitos casos a diferença de dias ou até meses entre um *check-in* e outro. Assim, apesar de haver milhões de check-ins no conjunto de dados fornecidos, em geral, haveria uma grande redução dessa dimensão quando os check-ins fossem transformados em prováveis ou possíveis viagens.

Por outro lado, a análise das informações referentes aos locais revelou que havia uma quantidade de locais sem uma categoria associada, ou uma categoria sem possibilidade de correspondência com um motivo da viagem realizada. Esse fato contribuiria para uma redução ainda maior na transformação do conjunto de chegadas em possíveis viagens, uma vez que um mesmo local poderia figurar em mais de uma viagem.

Sendo assim, foi elaborado um programa em linguagem computacional para refazer a coleta de dados, junto ao *Foursquare*, para a parte fornecida referente aos locais onde as chegadas são emitidas, refazendo parcialmente, portanto, a coleta realizada por quem nos forneceu tal conjunto de dados proveniente das redes sociais. Com vistas ao futuro uso neste estudo desses dados provenientes de redes sociais, para essa nova coleta foram identificados os locais sem categoria ou com categoria dúbia, como por exemplo, a categoria “Outros - Prédios”.

Essa atualização para categorias que não permitem direta associação a um motivo da viagem foi motivada tanto pela tentativa de repescagem da maior quantidade de locais

categorizados e hábeis a serem utilizados na correspondência com um atributo de viagem, o motivo, para a futura modelagem de escolha do modo de viagem, como pela oportunidade identificada de ter ocorrido uma nova versão de organização das categorias do *Foursquare*, por essa própria rede social, que poderia trazer contribuição para o uso desejado dessas categorias neste estudo – adiante neste texto descrita em “atributo motivo de viagem”.

Para cada cidade foi, então, selecionado um conjunto de locais sem categoria atribuída ou dúbia e realizada a nova coleta de categorias para esses locais através da *Foursquare API*¹. A tabela a seguir mostra a quantidade inicial de locais sem categoria e a quantidade remanescente posterior a nova coleta realizada, por exemplo, para algumas cidades.

Tabela 4.2 Quantidade de locais *Foursquare* recoletados, exemplo em algumas cidades

Recoleta de Informações dos Locais <i>Foursquare</i>				
	Total de locais	Total locais recoletados	Locais sem categoria antes da recoleta	Locais remanescentes sem categoria
Rio de Janeiro	25507	3942	2478	915
Paris	27627	3980	2766	888
São Paulo	65140	10445	5491	2089

Fonte: o próprio autor do texto.

O cruzamento entre as duas partes do conjunto de dados das redes sociais, as chegadas sequenciadas e tabuladas, e as categorias filtradas contendo a característica categoria, constituem a base de dados de deslocamentos intraurbanos, ou viagens, para a observação da mobilidade urbana em diversas cidades mundiais a ser empregada neste trabalho.

4.1.1 Filtragem dos dados

Esta filtragem dos dados provenientes das redes sociais procurou selecionar apenas sequências de chegadas que poderiam corresponder a uma real viagem daqueles usuários, bem como excluir dados considerados ruídos - *check-ins* sequenciados e espaçados por vários dias entre eles, usuários com apenas um *check-in*, etc.

¹ <https://developer.foursquare.com>

Todos os locais os quais remanesceram como sem categoria após a coleta realizada foram excluídos.

Todas as sequências de *check-ins* que não ocorressem no mesmo dia foram desprezados.

Para o enfoque desejado neste estudo, foram descartadas todas as viagens ocorridas no fim de semana, sábado e domingo.

A partir das sequências de *check-ins* ocorridos no mesmo dia, foram selecionados somente aqueles que possuíam uma diferença de tempo entre os *check-ins* maiores ou iguais a cinco minutos.

Com vistas a uma amostragem mais aderente à realidade, para aquelas sequências de *check-ins* ocorridos em dias diferentes, foi feita uma filtragem para repescagem daqueles onde a data do *check-in* de destino é o dia seguinte e sujeitas às restrições:

- a categoria *Foursquare* do *check-in* fosse relativa à moradia, como por exemplo “Casa – particular”, “Residência” e “Prédio Residencial – (Apartamento / Condomínio)”,
- a hora do *check-in* de origem fosse igual ou maior que dezoito horas,
- a diferença de tempo entre os *check-ins* fosse maior ou igual a cinco minutos.

Enfim, a filtragem reduziu, em geral, para menos da metade a quantidade de dados, possíveis viagens, disponíveis da base de dados das redes sociais. A tabela a seguir apresenta alguns exemplos resultantes dessa filtragem.

Tabela 4.3 Exemplos de redução na base de dados das redes sociais por filtragem

Redução base de dados redes sociais		
	Quantidade inicial sequencias de chegadas	Quantidade final de sequencias de chegadas após filtragem
Paris	193597	59531
Rio de Janeiro	374154	128275
São Paulo	1086003	423215
Tóquio	2648760	918883
Jacarta	3494758	1039817

Fonte: o próprio autor do texto.

4.1.2 Atributos de entrada

A base de dados das redes sociais fornecida não contém informações sobre as características socioeconômicas do viajante, comuns nas pesquisas origem-destino convencionais na área de Transportes, mas sim apenas as características das viagens realizadas, isto é, os atributos de viagem. Outra característica inerente a tal base de dados é a falta da informação sobre qual foi o modo de viagem utilizado para a realização do deslocamento intraurbano, como por exemplo carro, ônibus, metrô, trem ou caminhada.

Assim, para a realização do método proposto neste trabalho, foi necessário obter bases de dados disponíveis sobre viagens intraurbanas que contivessem a informação sobre o modo de viagem utilizado em cada viagem, para o ajuste de um modelo de escolha de viagem capaz de prever quais foram os modos de viagens utilizados nas viagens da base de dados de deslocamentos em diversas cidades proveniente das redes sociais. Ocorre que a construção e o uso de um modelo para predição do modo de viagem estão restritos aos atributos, ou às variáveis, ou, ainda, às características disponíveis na base de dados a ser aplicado tal modelo preditor – a base de dados proveniente das redes sociais.

As informações obtidas de forma direta disponível ou elaboradas de forma indireta, ambas a respeito de uma viagem, foram utilizadas como atributos de entrada para que um modelo de escolha previamente ajustado proviesse como saída qual foi o provável modo de transporte utilizado em uma viagem pertencente ao conjunto de dados das redes sociais. Para que a modelagem de escolha do modo de transporte obtivesse uma razoável performance preditiva, uma vez que não se dispunha dos atributos socioeconômicos tipicamente empregados nos modelos de escolha do modo de transporte, buscou-se o aproveitamento de toda a informação disponível nas bases de dados das redes sociais, ou seja, o aproveitamento de todas as características ou atributos pertencentes a uma viagem. Além das pesquisas origem-destino, foram utilizadas outras informações obtidas, como por exemplo as distâncias viajadas por modos de viagem alternativos, coletadas através da plataforma *Google Maps API*², como será apresentado oportunamente neste texto.

4.1.2.1 Atributos de Viagem

² <https://developers.google.com>

O conjunto de informações ou características relativas ao modo de transporte utilizado para a realização de uma viagem, como por exemplo, a distância da viagem por determinado modo de transporte, o tempo gasto para realizar a viagem, o custo de realização da viagem, o consumo de energia, entre outros, correspondem aos atributos de viagem.

Alguns atributos de viagem foram obtidos diretamente da base de dados das redes sociais como por exemplo, as coordenadas geográficas, latitude e longitude, dos locais de *check-ins* que correspondem aos pontos de origem e destino de uma viagem. Outros demandaram obtenção de forma indireta como por exemplo o atributo motivo de viagem o qual é obtido a partir da transformação da categoria *Foursquare* do local de destino da viagem. Ainda houve o caso da obtenção de atributos que demandaram uma nova coleta via *Internet* e por meio da plataforma *Google Maps API* como foi o caso das distâncias de viagem.

4.1.2.1.a Atributo motivo de viagem

No intuito de se obter uma boa performance preditiva do modelo de escolha do modo de viagem, o que demanda uma boa alimentação de informações sobre o domínio onde tal modelo será aplicado, foi utilizado como atributo de entrada a chamada categoria do lugar, conforme classificada pela rede social *Foursquare*. Esse atributo nominal é empregado por aquela rede social para uma classificação de todos os reais locais lá representados, por exemplo “Restaurante Brasileiro”, “Ponto de Hotdog”, “Universidade”, “Escritório”, “Parque Público”, “Escritório Corporativo”, “Residência”, etc – vide Figura 4.3.

Figura 4.2 Exemplo, vista dos tipos de locais de parte da cidade de Yokohama contida na base de dados das redes sociais

Fonte: o próprio autor do texto.

Porém, os dados das pesquisas origem-destino, de São Paulo e Londres, a serem empregados para o ajuste do modelo preditivo de escolha do modo de viagem dispõem de suas próprias e distintas entre si formas de classificação para o motivo da viagem realizada. Portanto, foi necessária a realização de um mapeamento por correspondência que transformasse as diferentes e numerosas categorias *Foursquare* nas poucas categorias de uma pesquisa origem-destino. Para essas últimas ainda foi realizada uma correspondência entre cada uma das pesquisas origem-destino para uma criada como padrão comum entre elas, sendo essa categoria criada e padronizada a que foi usada para a transformação das *Foursquare*. A tabela a seguir dispõe o mapeamento das categorias do atributo motivo de viagem das duas pesquisas origem-destino que se dispunha, de São Paulo e Londres, e um atributo de viagem padrão criado para fins deste estudo.

Tabela 4.4 Mapeamento entre as categorias dos atributos motivos de viagem das pesquisas origem-destino e um padrão a ser utilizado neste trabalho

Motivo de viagem - Londres*	Motivo de viagem - São Paulo	Motivo padrão
2 Work - Usual workplace	1 Trabalho/indústria	1 Trabalho
3 Work - Delivery/loading	2 Trabalho/comércio	
4 Work - Other	3 Trabalho/serviços	
1 Home	8 Residência	2 Residência
9 Hotel / holiday home		
8 Education (as a pupil)	4 Escola/educação	3 Educação
6 Shopping - Food	5 Compras	4 Compras
21 Shopping - Other		
5 Entertainment/recreation	7 Recreação/visitas/lazer	5 Lazer
17 Participate in Sport		
18 Leisure trip - enjoyment		
19 Visit friends/relatives at home		
15 Health or medical visit	6 Médico/dentista/saúde	6 Saúde
7 Personal business / use services	10 Assuntos pessoais	7 Outros
10 Drop off/pick up someone to/from work	9 Procurar emprego	
11 Drop off/pick up someone to/from school		
12 Drop off/pick up someone to/from health visit		
13 Worship or religious observance		
16 Drop off/pick up someone to/from other place		
14 Other		
20 Other Social		

Fonte: Pesquisa OD2007 de São Paulo, pesquisas LTDS 2007, 2008, 2009, 2010, e 2011 de Londres. Tabela e motivo padrão o próprio autor deste texto.

A partir da criação de um motivo de viagem padrão, apresentado na tabela acima, foi realizado o mapeamento de correspondência das mais de novecentas categorias *Foursquare* para as categorias desse motivo de viagem. A Tabela 4.6 apresenta, por exemplo, um pequeno trecho desse mapeamento.

Tabela 4.5 Trecho do mapeamento de correspondência entre as categorias *Foursquare* e o motivo padrão

Mapeamento categorias Foursquare X Motivo padrão	
Categorias <i>Foursquare</i>	Motivo padrão
Convenience Store	Compras
Conference Room	Trabalho
Confucian Temple	Outros
Convention Center	Trabalho
Apres Ski Bar	Lazer
Eye Doctor	Saúde
Home (private)	Residência
Hospital	Saúde
Hot Dog Joint	Compras

Fonte: o próprio autor do texto.

4.1.2.1.b Distâncias dos modos alternativos de viagem

Os atributos, ou variáveis, de entrada referentes às distâncias de viagens por diferentes modos alternativos para a realização das viagens foram coletados da plataforma *web Google Maps API*, por meio de software elaborado por este autor. Tal plataforma *web* possui os sistemas de transportes das cidades já implementados em seu sistema digital, que busca aproximação fiel do mundo real representado, considera a malha viária, bem como os horários das linhas de transporte público disponíveis, e fornece diariamente informações a quem as solicite, sobre as alternativas de rotas nos modos de transporte disponíveis em uma cidade para um desejado deslocamento intraurbano.

As distâncias coletadas para todas as viagens de uma cidade foram:

- distância total por modo carro;
- distância total por modo ônibus;
- distâncias de acesso e egresso ao modo ônibus;
- distância total de caminhada por modo ônibus, que se refere ao total caminhado incluindo trechos de caminhada entre aqueles a bordo de linhas de ônibus ou metrô, quando houve transbordo ou uso de mais de uma linha de ônibus ou metrô;
- distância a bordo do ônibus;

- distância a bordo do metrô quando houve trecho de metrô na viagem solicitada por modo ônibus;
- distância total por modo metrô;
- distâncias de acesso e egresso ao modo metrô;
- distância total de caminhada por modo metrô, que se refere ao total caminhado incluindo trechos de caminhada entre aqueles a bordo de linhas de metrô ou ônibus quando houve transbordo ou uso de mais de uma linha de metrô ou ônibus;
- distância a bordo do metrô;
- distância a bordo do ônibus quando houve trecho de metrô na viagem solicitada por modo metrô.

4.1.2.1.c Atributo duração da viagem

A base de dados das redes sociais dispõe dos horários de publicação das chegadas dos usuários aos locais, chamado de horário do *check-in*, o qual se supõe ser aquele da imediata chegada em algum local desejoso a informar de forma pública. Porém não há informação sobre o tempo de permanência no local ou, para o caso de viagens encadeadas, sobre o horário de saída que possibilite a estimativa do tempo efetivamente gasto para se realizar um deslocamento de um local ao outro. Mas, uma das vantagens da coleta das distâncias de viagem através da plataforma *Google Maps* é que informa o tempo gasto para realização da viagem em cada diferente modo de transporte alternativo em dado horário, e, portanto, foi um atributo coletado e utilizado de forma direta como atributo de entrada ao modelo de escolha. Assim, para todas as viagens de uma cidade foram coletados os atributos: durações das viagens por modos carro, ônibus e metrô. Tais atributos foram fornecidos pela plataforma *web* vinculados a outro atributo componente da solicitação pelas distâncias da viagem, o atributo hora no destino.

4.1.2.1.d Atributo hora no destino

Também de forma a fornecer auxílio na tarefa de predição do modo de viagem do modelo de escolha a ser ajustado, foi utilizada a hora de chegada no ponto de destino como

atributo de entrada para tal modelo. Esse atributo também foi utilizado para a coleta através do *Google Maps*, pois essa informação permite que a ferramenta *Google Maps* informe a distância ao modo de viagem solicitado mediante a disponibilidade de tal modo de transporte naquele horário, trazendo, assim, uma nuance real da situação de escolha do modo de viagem por um viajante e, esperançosamente, contribuinte à performance preditiva do modelo de escolha modal.

4.1.3 Tamanho das amostras de viagens das cidades

Para a determinação de um tamanho de amostra do conjunto de viagens disponível proveniente das redes sociais, foi utilizado um expediente simples, descartada a clássica forma estatística (SMITH, 1979) para esse tipo de amostragem, a qual foi empregada em estudo anterior deste autor (WERMERSCH, 2002) também sobre atributos de viagem em uma cidade e que, mesmo para um intervalo de confiança de 95% e erro de estimativa de 5%, determina um tamanho de amostra muito inferior – 780 viagens para o atributo tempo de viagem por modo caminhada – ao tamanho de amostra pretendido no presente estudo.

Considerando que há um limite diário de quantidade de coletas de 100.000 viagens, e que para algumas cidades, como por exemplo Paris e Toronto, somente se dispunha de aproximadamente 60.000 e 70.000 viagens ao todo, foi escolhido o tamanho de amostra de 100.000 viagens a serem coletadas as informações a respeito das cidades. Assim, se houvesse três modos de transporte disponíveis em uma cidade a ser realizada a coleta, a quantidade de dados a serem coletados seria de 300.000 coletas de distâncias de viagens, ou seja, o tamanho da amostra multiplicado pela quantidade de modos alternativos à realização das viagens da amostra. A amostragem foi realizada por sorteio aleatório dentre as viagens disponíveis após o processo de filtragem descrito no item 4.1 deste capítulo. Por fim, um programa em linguagem computacional foi elaborado e a coleta via *Internet* através da *Google Maps API* foi realizada para as cidades enfocadas neste estudo e os modos de viagem alternativos escolhidos para cada cidade, conforme adiante descrito neste texto.

4.1.4 Escolha das cidades a serem enfocadas neste estudo

A limitação de custo de coleta conduziu a uma redução na quantidade de cidades pretendidas ao enfoque deste estudo, custo esse à época da coleta de aproximadamente 50

dólares americanos para cada 100.000 pares origem-destino por um modo de viagem. Tendo, então, definido uma mesma quantidade de viagens a representar a mobilidade urbana em cada cidade, o critério do tamanho populacional em associação à disponibilidade de demais informações, como por exemplo os consumos específicos de todos modos de transportes da cidade e a disponibilidade da *Google Maps API* de provimento das distâncias por todos os modos alternativos escolhidos à coleta em determinada cidade.

Primeiramente um ordenamento populacional em termos de região metropolitana elencou algumas cidades e dez delas foram escolhidas para o enfoque deste estudo, vide Tabela 4.7.

Tabela 4.6 Cidades escolhidas a princípio

Cidades previamente escolhidas	
<i>Rank</i> populacional*	Cidade / Área urbana
1	Tokyo–Yokohama
2	Jakarta (Jabodetabek)
5	Seoul–Gyeonggi–Incheon
9	Nova Iorque
11	São Paulo
18	Los Angeles
26	Rio de Janeiro
29	Paris
32	Londres
53	Toronto

Fonte: o próprio autor do texto. * *World Atlas 2017* (www.worldatlas.com)

Entretanto, posteriormente, quando da coleta das distâncias de viagens foi detectado que para a cidade de Seoul a plataforma *Google Maps* não fornecia a distância viajada por modo carro embora fornecesse para os modos de transporte público. Desta forma, foi descartada a cidade de Seoul. Ainda durante a coleta das distâncias foi também detectado que para a cidade de Tóquio não eram fornecidas as distâncias de viagens através da plataforma *Google Maps API*, embora tais distâncias eram informadas através de acesso por navegador de *Internet* na página *Google Maps* para uma viagem. Diversas tentativas foram feitas ao longo de três meses sem sucesso e, então, a cidade de Tóquio também foi descartada.

Portanto, as oito cidades enfocadas neste estudo são:

Jakarta,

Nova Iorque,
 São Paulo,
 Los Angeles,
 Rio de Janeiro,
 Paris,
 Londres e
 Toronto.

Figura 4.3 Distribuição mundial das 8 cidades enfocadas neste estudo

Fonte: o próprio autor do texto.

No Apêndice A se encontram figuras das cidades enfocadas contendo feições do sistema de transporte e os pontos origem-destino da amostra de viagens provenientes das redes sociais, e conjunto total de viagens disponível após filtragem.

4.2 PESQUISAS ORIGEM-DESTINO

Os dados utilizados para a obtenção do modelo de escolha do modo de transporte provavelmente utilizado naquelas viagens ocorridas em diferentes cidades espalhadas pelo mundo, e coletadas através de redes sociais, são provenientes das pesquisas origem-destino das cidades de São Paulo e Londres, e correspondem às regiões metropolitanas dessas cidades.

Foram realizados pedidos formais às respectivas agências competentes de diversas cidades por fornecimento de suas pesquisas origem-destino para fins de uso neste estudo, como por exemplo, Paris, Milão, Barcelona, Madrid e Amsterdã, mas somente Londres retornou com a possibilidade de fornecimento gratuito da sua pesquisa origem-destino. Posteriormente, descobriu-se que o nível desejado de precisão de localização dos pontos de origem e de destino de viagens, o mesmo disponível na pesquisa o-d de São Paulo, não poderia ser fornecido fora do território britânico e foi, então, um dos motivos da realização do estágio naquele país - a obtenção da pesquisa o-d de Londres.

Durante o estágio no território britânico, a pedido deste autor, o supervisor estrangeiro do estágio forneceu um meio de contato em Singapura para obtenção da pesquisa origem-destino daquela cidade, mas infelizmente tal foi infrutífero. Essa busca por diversas pesquisas origem-destino se justifica pela intenção de estudar a mobilidade urbana da forma mais abrangente possível, em diversas cidades mundiais. Uma maior quantidade de informações sobre viagens realizadas no espaço intraurbano para a construção de um modelo de escolha do modo de viagem poderia, hipoteticamente, trazer a esse não apenas uma maior acurácia preditiva, mas também, talvez, uma aproximação universal, ou norma comum, de conduta humana diante da situação de escolha por um modo de transporte para realização de uma viagem. Todavia, foram encontradas disponíveis ao uso neste estudo apenas duas pesquisas o-d, São Paulo e Londres.

Essas pesquisas origem-destino contam com o método de coleta por entrevista domiciliar, onde para cada domicílio aleatoriamente sorteado são coletadas informações por um entrevistador a respeito das viagens realizadas no dia anterior pelos moradores daquele domicílio. Nas duas pesquisas, a informação sobre a real escolha do modo de transporte utilizado na realização de uma viagem é relatada pelos moradores do domicílio sorteado, e tal informação é conhecida como dado, ou informação, de preferência revelada.

A periodicidade de realização dessas pesquisas é diferente entre elas – a de São Paulo é realizada a cada dez anos e a de Londres é realizada anualmente. Com efeito, o tamanho da amostra coletada é diferente também: a pesquisa de São Paulo mais recente e disponível à época deste estudo, a do ano de 2007, contém aproximadamente 190.000 viagens, enquanto que a de Londres contém aproximadamente 40.000 viagens em cada edição anual. De forma a haver um equilíbrio entre as quantidades de viagens a serem modeladas à escolha do modo viagem, foram

solicitadas e recebidas 5 pesquisas o-d de Londres, referentes aos anos 2007, 2008, 2009, 2010 e 2011.

Há também diferença entre o sistema de coordenadas adotados nessas pesquisas para a localização dos pontos de origem e destino das viagens, sendo na pesquisa de São Paulo o sistema de projeção Universal Transversa de Mercator (UTM) e na de Londres, o Sistema de Referência Nacional Militar (OSNG), cada um com suas respectivas características e datum. Para que fosse possível a espacialização dos dados em mesmo mapa digital conjuntamente aos dados das redes sociais, os quais possuem localização em outro sistema de coordenadas, as geográficas globais latitude e longitude no modelo WGS84, foram realizadas as transformações dos sistemas de coordenadas das duas pesquisas origem destino para o sistema de coordenadas geográficas globais latitude e longitude no modelo WGS84.

4.2.1 Definição das Alternativas de Modos de Transporte

Existem diversos modos de viagem constantes nas pesquisas origem-destino com diversas frequências de ocorrência de viagens. Porém, a construção de um modelo de escolha do modo de viagem que incluísse todos os possíveis modos de viagens de um espaço urbano demandaria uma quantidade grande viagens incorridas em cada um dos modos desejados à inclusão no modelo.

Os modos mais frequentes nas duas pesquisas o-d, de São Paulo e Londres, são: carro, ônibus, metrô e caminhada. O modo trem possui discrepante menor quantidade de viagens que os demais modos e se fosse incluído na construção, ou ajuste, de um modelo de escolha conforme o pretendido neste estudo, levaria a uma redução da amostra para o caso de ser utilizada uma amostra balanceada entre modos de viagens. A figura a seguir apresenta um gráfico da divisão modal da pesquisa o-d de São Paulo para cinco modos de viagem.

Figura 4.4 Divisão modal da pesquisa OD2007 de São Paulo capital, para 5 modos de viagem

Fonte: o próprio autor do texto.

Por outro lado, a coleta das distâncias de viagens através da plataforma *Google Maps API* fornece as distâncias para todos os modos de transporte públicos necessários a uma viagem. Dessa forma, o cálculo da energia consumida numa viagem abarcaria a parcela devida ao modo trem em todas as viagens onde esse ocorresse conjuntamente aos demais modos de transporte público, ônibus e metrô. Assim, não estariam incluídas no cálculo total de energia consumida na mobilidade urbana de uma cidade apenas as viagens realizadas exclusivamente por modo trem.

Quatro razões nortearam a decisão pela não inclusão do modo trem na modelagem de escolha por modo de transporte deste estudo. São elas:

- a) o modo trem possui pequena frequência de ocorrência em viagens nas duas pesquisas o-d, de São Paulo e Londres;
- b) o consumo específico do modo trem é pequeno em relação aos demais modos;
- c) sua inclusão no modelo levaria a uma redução expressiva no tamanho de uma amostra balanceada entre modos de viagem se tal tipo de amostra for desejada a ser utilizada no ajuste do modelo de escolha;
- d) e, ainda, como cada modo alternativo de escolha demanda a coleta de todas as viagens da amostra de cada uma das 8 cidades, a inclusão de mais um modo alternativo causaria uma expressiva elevação do custo de coleta via *Google Maps API*.

Sendo assim, para as cidades onde existe o modo de transporte metrô, foram coletadas as distâncias para 3 modos de transporte alternativos: carro, ônibus e metrô. Para Jacarta, a qual não dispõe do modo metrô, foram coletadas distâncias para apenas 2 modos de transporte alternativos à realização de uma viagem: carro e ônibus.

Portanto, um modelo foi ajustado para 4 modos de viagem para ser utilizado nas cidades com disponibilidade do modo metrô, incluindo o modo caminhada, e outro modelo para Jacarta com apenas 3 modos de viagem, carro, ônibus e caminhada.

5 OBTENÇÃO DO CONSUMO DE ENERGIA DA MOBILIDADE URBANA NAS CIDADES – APRENDIZADO DE MÁQUINA

Este capítulo se refere à etapa de estimativa de um índice de consumo de energia devido à mobilidade observada em cada cidade, e utilizado para a obtenção do ordenamento de consumo de energia entre as cidades enfocadas, ordenamento esse que é um dos objetivos deste estudo. A análise de resultados aqui, limita-se à verificação de acurácia preditiva dos modelos de aprendizado de máquina, empregados para a atribuição de um modo de viagem a cada uma das viagens coletadas nas diferentes cidades.

5.1 INTRODUÇÃO À MODELAGEM DE ESCOLHA DO MODO DE VIAGEM

Para a obtenção do consumo de energia decorrente de uma viagem, faz-se necessário o conhecimento sobre qual meio de transporte foi utilizado. As viagens disponíveis a este trabalho, coletadas em diversas cidades por meio de redes sociais, não dispõem de informação sobre qual foi o modo de viagem utilizado. Sendo assim, um expediente computacional probabilístico foi empregado de forma a se atribuir um modo de viagem, carro, caminhada, ônibus ou metrô, a cada uma daquelas viagens.

Uma possível forma de atribuir um modo de viagem se dá por um **modelo de escolha do modo de viagem**, ajustado a uma amostra de viagens de uma determinada cidade, em que se tenha sabido o modo empregado em cada viagem. Dessa forma, o modelo ajustado é utilizado para projetar ou atribuir o modo de viagem àquelas viagens, provenientes de rede social, cujo modo é desconhecido. Como não há a informação do modo de viagem empregado em todas as diferentes cidades enfocadas neste trabalho, foram utilizadas as cidades de São Paulo e Londres para ajustar um modelo de escolha de viagem comum a ambas, com a utilização dos dados ou

viagens delas de forma conjunta, na tentativa de se obter um modelo com desempenho menos específico a cada uma dessas duas cidades.

Assim, um modelo de escolha do modo de viagem ajustado, com dados conjuntos das pesquisas origem-destino das cidades de São Paulo e Londres, foi empregado para atribuir o modo de viagem às viagens das demais cidades deste estudo. Tal translação geográfica ou espacial de aplicação de modelos de escolha do modo de viagem é objeto de estudos já há algum tempo, motivados pela custosa coleta das pesquisas origem-destino, geralmente restrita a grandes cidades – estudos dessa transferibilidade espacial entre áreas intraurbanas: KOPPELMAN & WILMOT(1982, 1985), KOPPELMAN et al.(1985), KOPPELMAN & PAS(1986), SANTOSO & TSUNOKAWA(2005); e entre áreas interurbanas: WATSON & WESTIN(1975), ATHERTON & BEN-AKIVA(1976), STOPHER & WILMOT(1979), TALVITIE & KIRSHNER(1978), KOPPELMAN et al.(1985), ABDELWAHAB(2001), KARASMAA(2007), SANTOSO & TSUNOKAWA (2010), MAMUN & SABBIR (2012).

Em geral, esses estudos apresentam a translação, ou transferência espacial de aplicação, dos modelos de escolha do modo de viagem ajustado em uma localidade para outra como algo possível. E alguns deles utilizam uma pequena amostra coletada da localidade a ser aplicado o modelo para uma melhoria da performance preditiva do modo escolhido nas viagens, antes desconhecido, daquela localidade. Quando tal transferência espacial é feita sem o auxílio de uma pequena amostra coletada na localidade de aplicação, como contribuição para o ajuste do modelo na direção de uma melhor performance preditiva, ela é comumente chamada de transferência ingênua¹. Adiante no texto será discutida a forma de transferência espacial do modelo de escolha do modo de viagem utilizada neste trabalho.

5.2 APRENDIZADO DE MÁQUINA

Previamente a atual era *big data*, desde a década de 1970 surgiram técnicas computacionais que buscavam soluções para problemas reais, utilizando a aquisição de conhecimento de especialistas codificado em regras lógicas. Desde então, a crescente complexidade dos problemas a serem tratados e volume de dados demandaram o desenvolvimento de ferramentas computacionais autônomas capazes de criar por si próprias uma hipótese quando aplicadas a um domínio, e sem a dependência de intervenção humana ou conhecimento prévio de especialistas. Assim, essas novas ferramentas computacionais

autônomas utilizam apenas os dados pertencentes a um domínio para a indução de uma hipótese ou função; e a apresentação ou utilização de mais dados para o mesmo domínio melhora a performance das ferramentas, demonstrando seu aprendizado sobre aquele domínio.

Diversas definições são encontradas na literatura sobre o que é Aprendizado de Máquina (AM) (MITCHELL, 1997; FLACH, 2012; MURPHY, 2012) que, em poucas palavras, é *res* computacional com capacidade de melhorar seu desempenho na realização de alguma tarefa por meio da experiência. A habilidade de processo de indução de hipótese associada a experiência passada, observação de dados, a aprendizagem, associam o AM à área do conhecimento de Inteligência Artificial (IA).

Os algoritmos de AM possuem diferentes formas de representação do conhecimento adquirido sobre o domínio aplicado, ou hipótese induzida, na aprendizagem, chamados de viés (*bias*) de representação. Por exemplo, o conhecimento adquirido por Redes Neurais Artificiais (RNA) é representado por um conjunto de números reais associados aos pesos das conexões sinápticas de sua rede de neurônios e os valores limiares, também números reais, de seus neurônios. Por sua vez, o conhecimento adquirido por Árvores de Decisão (AD) é representado por uma estrutura hierárquica em forma de árvore, com nós internos contendo um atributo e um valor limiar associado a ele e nós externos, as folhas, associados a classes de uma tarefa de classificação desejada no domínio aplicado.

Além disso, cada diferente algoritmo de AM possui uma diferente forma de realizar sua busca por uma hipótese no espaço de possíveis hipóteses do domínio onde ele é aplicado, ao realizar sua aprendizagem por meio dos dados de treinamento, seu melhor ajuste aos dados. Por exemplo, um determinado algoritmo de AD chamado ID3 (QUINLAN, 1979) busca a obtenção de uma árvore de decisão com poucos nós. Tal forma é chamada de viés de busca. Os vieses, de busca e representação, restringem as hipóteses a serem visitadas no espaço de busca e são características inerentes a cada algoritmo de AM.

Diversos estudos empregaram algoritmos de AM para a modelagem da escolha do modo de viagem em função da melhor performance preditiva, em comparação a tradicionais modelos estatísticos - por exemplo o modelo *Logit* -, e apresentam os algoritmos de AM com a melhor performance observada (NIJKAMP et al., 1996; GONZÁLEZ-ESCRIVA & SANCHEZ, 2000; DAPENA & MEDINA, 2000; XIE et al., 2003; CANTARELLA & DE LUCA, 2005; MOONS et al., 2007; TORTUM et al., 2009, OMRANI et al., 2013).

Outros estudos comparam a performance de diversos algoritmos de AM, como CARUNA & NICULESCU-MIZIL (2006), WILLIAMS et al. (2006) e QIAN et al. (2015). Neste estudo, para a obtenção de um melhor modelo de escolha por modo de viagem, em termos de sua performance preditiva para a classificação de modos de viagem, foram investigados diferentes algoritmos de AM devido aos seus diferentes vieses relatados com boa performance preditiva à tarefa de classificação.

Uma breve descrição de cada algoritmo investigado para a tarefa de classificação almejada neste trabalho é apresentada a seguir. Entretanto, foge ao foco deste trabalho um detalhamento da teoria ou implementação prática de cada um desses algoritmos de AM, uma vez que há extensiva literatura na área de conhecimento pertinente ao assunto.

5.2.1 Árvores de Decisão

Os algoritmos do tipo Árvores de Decisão (AD) são uma estrutura hierárquica que abrange todo conjunto de dados, ou instâncias, do treinamento, construída a partir de um primeiro nó divisão, o nó raiz, com dois ou mais nós sucessores, os nós intermediários, e assim por diante, até o atingimento dos nós folhas que representam as classes. O nó raiz, bem como os nós intermediários, nós de divisão, contém um teste condicional baseado nos valores de um atributo que divide o espaço de dados de treinamento. O teste condicional envolve um ou mais atributos, um operador matemático (por exemplo $=$, $\geq$ etc) e um valor do domínio do atributo. Assim, a estratégia de construção da árvore de decisão é, em termos gerais, criar um nó adicional que dividirá o espaço de instâncias de treinamento, de tal forma a melhorar a performance preditiva de toda a estrutura da AD. Alguns algoritmos de AD ainda têm implementado uma rotina de poda da AD, para melhorar a habilidade de generalização a novos dados, os dados de validação do treinamento.

Foram utilizados três modelos de árvore de decisão na investigação inicial por melhores capacidades preditivas, o REPTree (HALL et al., 2009), o C4.5 (QUINLAN, 1993) e Random Forest (BREIMAN, 2001).

O algoritmo **REPTree** (HALL et al., 2009) é um algoritmo simples de AD que constrói de forma rápida sua árvore usando informação sobre o ganho por variância, e faz poda da árvore construída em face do desempenho preditivo.

O algoritmo **C4.5** (QUINLAN, 1993) usa uma estratégia mais complexa. Associa-se uma probabilidade a cada um dos possíveis valores do atributo. As probabilidades são estimadas com base nas frequências observadas dos valores para o atributo nos exemplos do nó corrente. Esses valores fracionários são usados para calcular o ganho de informação desse atributo. Quando se classifica uma instância de teste, o algoritmo passa a instância através de todos os ramos em que o valor do atributo desconhecido foi detectado. Cada ramo produz como saída um voto para a classe. A saída final é calculada como a classe majoritária de todas as saídas dos ramos.

O algoritmo *Random Forest* (**RF**), ou Floresta Aleatória, (BREIMAN, 2001), como sugere o nome, é um conjunto de árvores de decisão, a floresta, onde cada árvore é construída a partir de partição aleatória do conjunto de atributos de entrada de um domínio a classificar, bem como de uma partição aleatória das instâncias de treinamento. Como observa FLACH (2012)], em analogia a uma AD onde, em última análise, suas folhas particionam o espaço de instâncias a classificar, a correspondente partição do espaço de instâncias a classificar por uma Random Forest é essencialmente a interseção das partições das árvores individuais no conjunto de árvores, floresta. Ao estimar um resultado mediando de forma conjunta a partir de diversos modelos de AM, no caso de diversas AD, RF obtém redução da variância de uma estimativa - deficiência frequentemente atribuída a uma AD (MURPHY, 2012). Embora neste texto esse algoritmo esteja descrito no subitem sobre AD, ele é usualmente classificado na literatura distintamente das AD, como um algoritmo do tipo “conjunto de modelos” (FLACH, 2012).

5.2.2 Redes Neurais Artificiais

Existem vários tipos (modelos) de Redes Neurais Artificiais (RNA), com indicação própria a cada diferente tarefa. O tipo de RNA investigado neste trabalho para a tarefa de classificação almejada é conhecido como *feed-foward multilayer perceptron* (**MLP**) (HAYKIN, 1999). À semelhança do modelo biológico, uma RNA é composta de unidades de processamento (“neurônios”), contendo valores limiares de ativação e uma função de transferência a seguinte camada, dispostos em camadas e interligadas por conexões (“sinapses”), as quais são associadas a pesos que ponderam a entrada recebida por cada unidade da rede. A camada de entrada recebe os valores dos atributos de entrada referentes a uma instância de treinamento, que alimenta adiante, através das conexões ponderadas por pesos, as unidades de processamento da(s) camada(s) intermediária(s), e então as unidades da camada de

saída. A saída calculada é então comparada com a saída desejada, no caso, o modo de viagem realizado naquela instância (viagem) de treinamento, e o erro é propagado de volta em direção à camada de entrada para o ajuste dos parâmetros da rede, dos pesos sinápticos e dos valores limiares de ativação das unidades de processamento.

O conhecimento adquirido (aprendido) sobre o domínio pela rede MLP está, portanto, armazenado na forma com que as unidades de processamento estão conectadas em sua arquitetura e em seus parâmetros ajustáveis. Nos experimentos de AM realizados neste trabalho, uma rede MLP é nomeada de acordo com a disposição das unidades de processamento nas camadas, ou seja, de acordo com sua arquitetura. Por exemplo, a rede 25-30-4 significa que possui 25 unidades de entrada, 30 unidades intermediárias e 4 unidades de saída. As unidades de entrada correspondem aos atributos de entrada e as unidades de saída às classes a classificar, no caso, os modos de viagem: carro, ônibus, metrô e caminhada.

5.2.3 Máquinas de Vetor de Suporte – Support Vector Machines

Em poucas palavras, o algoritmo Support Vector Machines (**SVM**) (VAPNIK et al., 1997), em uma tarefa de classificação, busca encontrar um hiperplano que separe as classes, do domínio em questão, localizado com a maior margem possível entre as instâncias de treinamento de diferentes classes mais próximas entre si. Por exemplo, para um domínio com somente dois atributos de entrada, espaço bidimensional, e duas classes a classificar, o hiperplano pode ser simplificado por uma reta disposta de tal forma otimizada com a maior margem entre algumas instâncias de treinamento das duas diferentes classes, as quais se localizam mais próximas entre si. Tais instâncias, que dão suporte à localização otimizada da linha divisória entre as classes, são chamadas de vetores de suporte, daí o nome do algoritmo. Quando a dificuldade de separação de classes aumenta, o algoritmo projeta as instâncias de treinamento em um espaço dimensional maior que o do problema original a classificar, na esperança de encontrar um hiperplano de separação entre as classes nesse espaço dimensional maior. Em geral, a projeção em espaço dimensional aumentado é feita através de uma função polinomial ou de base radial. Tais funções são chamadas nesse algoritmo como funções *kernel* e são parâmetros de escolha ao treinamento do algoritmo.

5.3 EXPERIMENTOS DE APRENDIZADO DE MÁQUINA

Os experimentos para obtenção do desempenho preditivo dos algoritmos de AM foram conduzidos com aferição do desempenho preditivo, utilizando amostragem por validação cruzada, também conhecida como *k-fold cross-validation* (MURPHY, 2012). Nesse método de amostragem por validação cruzada, o conjunto de instâncias disponíveis é dividido aleatoriamente em k subconjuntos aproximadamente iguais e as instâncias de $k - 1$ partições são utilizadas no treinamento do algoritmo preditor, o qual é, então, testado com uso da partição restante.

Em geral, os treinamentos foram realizados com 10 partições do conjunto de instâncias disponível, ou seja, *10-fold cross-validation*, e repetido cada evento de treinamento 3 vezes, de forma a se obter uma estimativa da medida de desempenho de um algoritmo no domínio em questão, com um desvio padrão associado a esses três eventos de 10 treinamentos cada, totalizando 30 treinamentos para cada algoritmo de AM investigado.

A medida de desempenho adotada em todos os experimentos conduzidos foi a média de classificações corretas na predição do algoritmo treinado, obtida através das partições de validação, ou seja, a performance média de todos os k -modelos treinados (ajustados). Os experimentos iniciais, investigativos, apresentam somente um percentual total médio abrangendo as quatro classes, os modos de viagem, corretamente classificados. Após um algoritmo ter sido escolhido como o preditor do modo de viagem para o presente trabalho, será realizado um treinamento com a apresentação tanto da medida de desempenho geral a todas as classes, bem como o desempenho individual a cada uma delas.

O conjunto de dados empregado no treinamento é aquele constituído de todas as instâncias, no caso, viagens, disponíveis após o processo de limpeza de dados, oportunamente descrito neste texto, das duas pesquisas origem-destino disponíveis, São Paulo e Londres. Esse conjunto foi, então, balanceado em termos das classes, os modos de viagem, sendo tomado o modo de viagem de menor quantidade de instâncias, viagens por metrô em Londres com 10.473 instâncias, como o limite de corte para os demais modos em ambas as cidades. Desta forma, obteve-se um conjunto de dados disponível contendo 83.784 viagens, rotuladas em 4 classes, quer sejam os modos carro, ônibus, metrô e caminhada, balanceado por classe e quantidade de instâncias para ambas as cidades.

5.3.1 Resultados dos Experimentos de AM

Todos experimentos realizados neste estudo com AM utilizaram todos os atributos, ou variáveis, de entrada descritos no item 4.1.2 acima neste texto. Devido ao fato de o atributo de entrada referente ao motivo de viagem ter sido transformado em variável *dummy* de sete posições, os experimentos envolvem vinte e quatro variáveis de entrada e uma variável de saída referente ao modo de viagem.

O primeiro experimento realizado envolveu quatro diferentes algoritmos de AM, são eles: o REPTree, o C4.5 (implementado com o nome de J48 no software utilizado), *Random Forest*, e um modelo simples de redes neurais artificiais MLP de apenas uma camada intermediária de dois neurônios (implementado com o nome de *MLPClassifier*). Nesse primeiro experimento investigativo sobre a performance dos algoritmos, todos eles foram utilizados em seus parâmetros padrão conforme implementados no software de AM utilizado (HALL et al., 2009). A figura a seguir apresenta o relatório de saída do software utilizado contendo os resultados desse primeiro experimento.

Figura 5.1 Relatório de resultados do 1º experimento de diferentes algoritmos de AM

Dataset	(1) trees.REPTree	(2) trees.J48	(3) trees.Rando	(4) functions.M
'sp_LD_4mods_ba(30)	62.77(0.53)	60.61(0.48) *	66.92(0.53) v	61.20(0.52) *

Fonte: o próprio autor do texto.

O algoritmo de performance vencedora neste primeiro experimento foi o *Random Forest*, com o desempenho médio de classificação correta entre as classes de 66,9% e desvio padrão de 0.53%.

Para os algoritmos SVM, foram realizados dois experimentos separadamente devido ao tempo de processamento e a capacidade de memória demandados. Um experimento exclusivo para o algoritmo SVM foi realizado primeiramente com a função *kernel* padrão, tipo polinomial, implementada no algoritmo, com a mesma validação cruzada *10-fold* do experimento realizado anteriormente com os demais algoritmos de AM da Figura 5.1 acima. A Figura 5.2 apresenta o resultado desse primeiro experimento com o SVM.

Figura 5.2 Relatório de resultado do 1º experimento de SVM

Dataset	(1) functions.SMO
'sp_LD_4mods_baleq_full_n (30)	62.78(0.43)

Fonte: o próprio autor do texto.

Noutro experimento conduzido com o algoritmo SVM, a função kernel utilizada foi a função de base radial, opcionalmente implementada no algoritmo SVM do software utilizado, que, devido ao tamanho do conjunto de instâncias e a demanda por recursos computacionais pela forma de processamento desse algoritmo, em vez de ser realizada a validação cruzada 10-*fold*, foi realizada uma validação cruzada através de conjunto de validação composto por sorteio aleatório de 30% das instâncias do conjunto total, e repetido o processo quatro vezes, à semelhança do proposto por PRECHELT (1994). O relatório da Figura 5.3 apresenta o resultado desse segundo experimento de SVM.

Figura 5.3 Relatório de resultado do 2º experimento de SVM

Dataset	(1) functions.SMO	'-

'sp_LD_4mods_baleq_full_n	(4)	58.02 (0.44)

Fonte: o próprio autor do texto.

Devido à baixa performance alcançada nos dois modelos de SVM, ajustados com duas diferentes funções, comparada à performance do algoritmo RF do primeiro experimento entre diferentes algoritmos, a opção pelo uso do algoritmo SVM neste trabalho foi descartada.

Ainda em busca de melhores performances dos algoritmos além daquela do primeiro experimento, a investigação foi ampliada para o algoritmo de redes neurais artificiais para mais cinco diferentes arquiteturas:

Algumas com uma camada intermediária de:

- 8 unidades de processamento intermediárias;
- 14 unidades de processamento intermediárias;
- 16 unidades de processamento intermediárias;
- 30 unidades de processamento intermediárias; e
- uma com duas camadas intermediárias de processamento, uma delas com 20 e outra com 8 unidades de processamento.

O gráfico da Figura 5.4 apresenta os resultados de performance das RNA MLP investigadas.

Figura 5.4 Gráfico dos resultados das Redes Neurais Artificiais MLP

Fonte: o próprio autor do texto.

Os resultados das RNA mostram uma estabilidade da performance preditiva dos diferentes modelos abaixo dos 65% de acurácia média preditiva. Mesmo diversificando a arquitetura dos modelos, nenhum dos experimentos, de 30 treinamentos cada, realizados com as RNA, alcançou o resultado do primeiro experimento com o algoritmo *Random Forest*, ainda que esse primeiro modelo de RF fosse o mais singelo com apenas 100 árvores, conforme parâmetro padrão pré-estabelecido do software utilizado.

Então, um experimento com mais modelos de RF com variação do parâmetro de quantidade de árvores a serem construídas na floresta foi realizado para 200, 300 e 500 árvores. O gráfico da Figura 5.5 apresenta os resultados contidos no relatório, Figura 5.6, de saída do processamento de 30 treinamentos para cada modelo de RF, incluindo aquele do primeiro experimento com 100 árvores apenas.

Figura 5.5 Gráfico dos resultados dos algoritmos Random Forest

Fonte: o próprio autor do texto.

Figura 5.6 Relatório de resultado de performance dos algoritmos RF

Dataset	(1) trees.RandomFo	(2) trees.Rando	(3) trees.Rando
'sp_LD_4mods_bal (30)	67.10 (0.50)	67.17 (0.44)	67.23 (0.44)

Fonte: o próprio autor do texto.

Os resultados do gráfico acima mostram as performances com tendência assintótica um pouco inferior ao valor de 67,5%, mesmo que a quantidade de árvores seja mais que dobrada na floresta, com variação de apenas algumas unidades na segunda casa decimal do percentual médio de acerto entre as classes no caso das duas florestas maiores, e acomodação do desvio padrão em 0.44% nas florestas maiores, a de 300 e a de 500 árvores. Portanto, para este trabalho, foi escolhido o modelo de RF de 500 árvores, como o modelo preditor para a tarefa de escolha do modo de viagem para os conjuntos de instâncias das cidades donde não se tem qual foi o modo escolhido naquelas viagens realizadas. A Figura 5.7 apresenta a performance do modelo classificador RF com 500 árvores e quatro classes.

Figura 5.7 Relatório de resultado de performance do modelo classificador RF com 500 árvores e quatro classes

a	b	c	d	<-- classified as	TP Rate
8526	5822	3170	3428	a = CAR	0.407
4239	11843	3489	1375	b = BUS	0.565
767	1683	18298	198	c = SUBWAY	0.874
1669	1345	341	17591	d = WALK	0.840

Fonte: o próprio autor do texto.

5.3.2 Investigação sobre a transferência espacial do modelo preditor de escolha do modo de viagem

A literatura, acima citada, apresenta que a agregação de informação sobre a localidade de aplicação traz uma melhora preditiva na aplicação do modelo em comparação com a transferência direta do modelo, chamada de ingênua. Tal agregação pode se dar de diversas formas; geralmente se utiliza uma pequena amostra coletada na localidade de aplicação para proceder uma atualização (dos parâmetros) do modelo ajustado. Outra forma de agregação de informação é a inclusão de uma variável destinada a informar uma variação do ambiente externo ao modelo, como por exemplo a divisão modal.

Para o caso deste trabalho, não há a possibilidade de se lançar mão a uma pequena coleta de viagens nas localidades de aplicação que contenha a informação sobre qual foi o modo de viagem escolhido nas viagens realizadas.

Desta forma, para uma verificação a respeito da transferência geográfica para outras cidades do modelo preditor de escolha do modo de viagem, ajustado para as duas cidades Londres e São Paulo, foi utilizada a informação sobre a divisão modal para um teste comparativo contra um modelo preditor ajustado com balanceamento entre as classes, ou seja, contra um modelo que não observou diferença entre as quantidades de viagens atribuídas a cada modo – todos os modos com mesma quantidade de instâncias/viagens.

A questão de conjunto de dados desbalanceados, ou seja, dados pertencentes à uma ou mais classes que aparecem com uma frequência maior que os dados das demais classes no conjunto, é tratada na literatura sobre AM como um problema a ser sanado, pois é sabido que vários algoritmos de AM, quando alimentados com dados desbalanceados na fase de aprendizagem (treinamento), posteriormente quando usados para a predição de novos dados,

tendem a privilegiar a classificação em favor da(s) classe(s) majoritária(s), ou seja, aprendem sobre a distribuição (desbalanceamento) das classes também.

Essa aprendizagem sobre o desbalanceamento entre classes (modos de viagem) das instâncias de treinamento poderia ser, então, utilizada como uma passagem de informação a respeito do domínio de aplicação transladado, a localidade diferente de aplicação. Em hipótese, tal figuraria, portanto, como uma atualização - termo utilizado na literatura acima citada - sobre a transferência espacial de modelos de escolha do modo de viagem. Ou ainda, essa aprendizagem sobre o desbalanceamento adiciona um viés a respeito da localidade onde será aplicado o modelo preditor.

Procedeu-se, então, com a criação de duas amostras, dois conjuntos de instâncias de viagens, um para São Paulo e outro para Londres, que embora contenham conjuntamente viagens de ambas cidades, possuem um desbalanceamento, divisão modal, conforme os percentuais observados nas duas pesquisas-origem de cada uma delas respectivamente.

A criação dessas amostras foi feita de forma que o número total de instâncias, viagens, fosse igual àquele do conjunto balanceado utilizado nos experimentos de AM, 83.784 viagens. Isso se dá de forma que a comparação de performances preditivas seja estipulada sobre uma mesma porção de conhecimento extraído do domínio em questão. Ainda, em aproximação a uma condição *coeteris paribus*, a composição das instâncias dessas duas amostras partiu do conjunto de classes balanceadas, com equalização das quantidades por redução ou adição de novas instâncias, quando necessário.

A adição de novas instâncias foi feita por sorteio aleatório no conjunto proveniente das pesquisas origem-destino, mantidas as primeiras instâncias adicionadas a cada classe, e incrementado por outras novas instâncias para o atingimento das quantidades necessárias conforme o desbalanceamento almejado.

Essa composição que mantém as instâncias já incluídas em uma das duas amostras para a composição da outra, e somente adiciona novas instâncias à classe necessária a ser aumentada, tem vistas a que seja oportunizado ao aprendizado de um algoritmo apenas o desbalanceamento entre as classes almejado, e não diferente informação.

Já para fins de treinamento dos algoritmos de AM, essas duas amostras foram divididas em uma partição contendo 70% dos dados para o treinamento do algoritmo RF escolhido e 30%

deixadas para teste. O mesmo foi feito para um novo treinamento com o conjunto de dados balanceados, anteriormente utilizado nos experimentos acima descritos, agora dividido em uma partição de treinamento com 70% dos dados e uma partição de 30% para teste. Obteve-se, portanto, três conjuntos de dados mistos de São Paulo e Londres para a realização de treinamentos e testes. Bem como cuidou-se para que as três partições de teste contivessem dados exclusivos em relação àqueles utilizados aos treinamentos dos três algoritmos RF. As Figuras 5.8 e 5.9 mostram as amostras em suas classes desbalanceadas.

Figura 5.8 Amostra com iguais quantidade de viagens de São Paulo e Londres desbalanceadas conforme a divisão modal da pesquisa origem destino de São Paulo

Fonte: o próprio autor do texto.

Figura 5.9 Amostra com iguais quantidades de viagens de São Paulo e Londres desbalanceadas conforme a divisão modal da pesquisa origem destino de Londres

Fonte: o próprio autor do texto.

As duas partições de teste desbalanceadas conforme a divisão modal de São Paulo e Londres foram utilizadas para verificação do percentual médio de acerto à classificação dos três algoritmos treinados: um algoritmo treinado com balanceamento de classes, e outros dois algoritmos treinados com desbalanceamento de classes conforme São Paulo e Londres.

O intuito dessa verificação por teste cruzado foi simular a situação quando o algoritmo RF treinado com o balanceamento de classes for projetar os modos àquelas viagens de outra localidade, a transferência espacial ingênua, e comparar com a projeção dos algoritmos treinados com o desbalanceamento de classes, isto é, informados sobre a localidade de projeção, uma vez que a literatura aponta haver ganho de performance preditiva quando há a passagem de alguma informação ao modelo preditor. A Figura 5.10 apresenta os resultados do treinamento dos algoritmos RF e a Tabela 5.1 apresenta os resultados dos testes cruzados.

Figura 5.10 Gráfico dos resultados dos algoritmos Random Fores treinados com 70% dos dados disponíveis e com desbalanceamento de classes conforme a divisão modal de Londres e São Paulo

Fonte: o próprio autor do texto.

Tabela 5.1 Resultados da aplicação das partições de testes, contendo 30% das instâncias, aos algoritmos RF treinados com as respectivas partições de treinamento, contendo 70% das instâncias

Percentuais Médios de Instâncias Corretamente Classificadas			
Modelos	Amostras		
	Teste 30% classes balanceadas	Teste 30% divisão modal Londres	Teste 30% divisão modal S.Paulo
RF treinado classes balanceadas	66,83	63,02	64,06
RF treinado divisão modal Londres	---	69,40	---
RF treinado divisão modal S.Paulo	---	---	68,67

Fonte: o próprio autor do texto.

Apesar de não se ter trinta ou mais pesquisas origem-destino para serem testadas, com vistas a um melhor suporte estatístico, mas apenas duas delas, os resultados da tabela acima se coadunam com a hipótese de que o desbalanceamento das classes do conjunto de treinamento conforme a divisão modal da localidade de aplicação, aumenta a acurácia preditiva do modelo preditor para aquela localidade:

- 63, 2% contra **69,40%** para Londres e
- 64,06% contra **68,67%** para São Paulo.

Nota-se que a performance preditiva do modelo RF treinado com classes balanceadas foi inferior ao ser testado com dados com desbalanceamento de classes, ou seja, dados com distribuição de classes conforme a divisão modal da localidade de aplicação.

Pode-se também notar a partir das matrizes de classificação de cada modo, extraídas dos relatórios de testes dos três algoritmos, que houve uma melhora significativa de performance preditiva para a classe carro, também seguida por uma piora significativa na classe ônibus, vide figuras a seguir.

Figura 5.11 Teste de 30% das instâncias contra RF treinado com 70% das instâncias ambos os conjuntos, teste e treinamento, com classes balanceadas

a	b	c	d	<-- classified as	TP Rate
2556	1770	934	1024	a = CAR	0.407
1274	3502	1078	430	b = BUS	0.557
219	502	5474	89	c = SUBWAY	0.871
486	425	108	5265	d = WALK	0.838

Fonte: o próprio autor do texto.

Figura 5.12 Teste de 30% das instâncias contra RF treinado com 70% das instâncias, ambos os conjuntos, teste e treinamento, conforme divisão modal de Londres

a	b	c	d	<-- classified as	TP Rate
6757	499	414	1480	a = CAR	0.738
2466	1024	307	391	b = BUS	0.245
695	141	1292	48	c = SUBWAY	0.594
1065	137	48	8368	d = WALK	0.870

Fonte: o próprio autor do texto.

Figura 5.13 Teste de 30% das instâncias contra RF treinado com 70% das instâncias, ambos os conjuntos, teste e treinamento, conforme divisão modal de São Paulo

a	b	c	d	<-- classified as	TP Rate
5881	743	463	1415	a = CAR	0.692
2348	1535	387	426	b = BUS	0.327
616	221	1571	56	c = SUBWAY	0.638
964	184	51	8271	d = WALK	0.873

Fonte: o próprio autor do texto.

Devido ao fato de os percentuais médios de instâncias corretamente classificadas nos testes pelos preditores treinados com desbalanceamento de classes terem sido superiores àqueles do preditor treinado com classes balanceadas, decidiu-se empregar o modelo de RF treinado com desbalanceamento de classes para a tarefa de atribuição do modo de viagem aos dados provenientes das redes sociais neste estudo.

A Tabela 5.2 apresenta os percentuais para a divisão modal em cada cidade a serem utilizados para a elaboração do desbalanceamento do conjunto de treinamento. Assim, da mesma forma como o realizado no experimento anterior para São Paulo e Londres, o desbalanceamento foi realizado a partir de inclusão de novas instâncias apenas na quantidade necessária para a equalização do conjunto de treinamento conforme a divisão modal da cidade a ser aplicado o algoritmo RF, sendo as demais instâncias comuns a todas as cidades - a minimização do efeito de ganho de conhecimento por novas instâncias não vistas por algoritmos treinados para outras cidades.

Tabela 5.2 Percentuais para o desbalanceamento da divisão modal

Percentuais Divisão Modal					
	Carro motor	Transporte	Ônibus	Metrô	Caminhada
Jacarta	0.322		0.39		0.288
Nova Iorque	0.144		0.088	0.173	0.595
São Paulo	0.339		0.187	0.097	0.378
Los Angeles	0.548		0.1	0.015	0.337
Rio de Janeiro	0.139		0.462	0.036	0.363
Paris	0.169		0.138	0.161	0.532
Londres	0.364		0.167	0.087	0.383
Toronto	0.617	0.298			0.085

Fonte: o próprio autor do texto conforme as seguintes fontes:

SUSILO et al. (2007);

Census 2010 U.S. Census Bureau, National Household Travel Survey 2009, New York City;

California Household Travel Survey 2013;

PDTU - Plano Diretor de Transporte Urbano da Região Metropolitana do Rio de Janeiro 2013;

Enquete Globale de Transport EGT 2010 - DRIEA Ile-de-France;

2006 Transportation Tomorrow Survey.

Tabela 5.3 Resultados de performance dos algoritmos RF preditores do modo de viagem por cidade

Percentual médio de acerto à classificação dos modos de viagem, algoritmo RF	
	%
Jacarta	67.4
Nova Iorque	80.5
São Paulo	68.7
Los Angeles	76.8
Rio de Janeiro	76.7
Paris	76.4
Londres	69.7
Toronto	71.8

Fonte: o próprio autor do texto.

Para a cidade de Paris foram estimados percentuais da divisão modal proporcional a quantidades de viagens contidas na área central e a coroa externa, região metropolitana, devido ao fato de a referência consultada apresentar diferentes divisões modais para essas duas áreas.

5.4 CÁLCULO DO ÍNDICE DE CONSUMO DE ENERGIA

Para o cálculo do índice de consumo de energia devido à mobilidade urbana observada nas diferentes cidades serão utilizados:

- O modelo classificador RF com 500 árvores previamente treinado com a amostra desbalanceada proveniente das pesquisas origem-destino das cidades de Londres e São Paulo,
- As amostras das 8 cidades coletadas por redes sociais e ferramenta *Google Maps* e contendo os atributos das viagens,
- Informações específicas sobre consumo de energia dos diferentes modos de viagem existentes em cada uma das 8 cidades, conforme UITP (2001).

Para a contabilização do consumo de energia para cada viagem da amostra de uma cidade, inicialmente será feita a atribuição do modo de viagem por meio do modelo classificador RF. A partir da definição do modo de viagem àquela viagem, todas as distâncias, em quilômetros, dos demais modos motorizados utilizados na viagem, serão utilizadas no cálculo do consumo total da viagem, através do simples produto entre as distâncias viajadas em cada modo e o consumo específico de cada modo de viagem utilizado respectivamente, dado

em MJ/quilômetro, de forma a se obter o valor da energia consumida na viagem, em MJ. Por exemplo, uma viagem à qual foi atribuído o modo ônibus e a partir da coleta das reais distâncias, através do *Google Maps*, foi coletado que a realização daquela viagem por modo ônibus também utiliza, além de caminhadas de acesso, egresso e intermediárias, os modos trem e metrô, todos os modos motorizados serão computados para a o consumo de energia dessa viagem.

Essa forma de cálculo a partir de aproximação da real forma na qual uma viagem pode ser realizada, conforme informado pelo *Google Maps*, para fins de obtenção do consumo de energia de uma viagem, funciona como se a atribuição do modo de viagem pelo modelo RF preditor fosse em alguma medida assistida no que diz respeito aos modos de transporte público. Por exemplo, se uma viagem é erroneamente atribuída ao modo ônibus e a coleta das distâncias do *Google Maps* informa que tal viagem possui distância zero para o ônibus, ou seja, não há como realiza-la por ônibus, ou não de uma forma razoável conforme o algoritmo *Google Maps* foi implementado, então o consumo de energia da viagem será calculado como zero.

A tabela a seguir apresenta, por exemplo, o consumo de energia específico a cada modo da cidade de Paris conforme disposto em UITP (2001).

Tabela 5.4 Exemplo: Consumo de energia por modo de viagem de Paris

Indicadores de Consumo de Energia por Modo de Viagem - Paris		
Consumo de energia por veículo de transporte privado	3,5 MJ/km	
Consumo de energia por veículo de transporte público por passageiro:		
Ônibus	1,59 MJ/p.km	
Tram	---	
Trem Leve	0,34 MJ/p.km	
Metrô	0,34 MJ/p.km	
Trem Suburbano	0,45 MJ/p.km	
Ferry	---	

Fonte: UITP (2001)

De forma a se obter um comparativo entre cidades sem a influência da característica específica de consumo unitário dos modos de transporte de cada localidade, foi realizada uma uniformização para um valor médio dos consumos unitários constantes em UITP (2001) e tal valor médio foi o utilizado no cálculo do indicador de consumo de viagem para todas as cidades enfocadas.

Tabela 5.5 Consumos médios unitários de energia para os modos de viagem a serem utilizados no cálculo do indicador de consumo de viagem devido à mobilidade urbana.

Indicadores de Consumo de Energia Médios - Para as oito cidades	
Carro motorista	4.06 MJ/km
Ônibus	1.25 MJ/p.km
Tram	0.41 MJ/p.km
Trem leve	0.52 MJ/p.km
Metrô	0.46 MJ/p.km
Trem suburbano	0.59 MJ/p.km
Ferry	2.20 MJ/p.km

Fonte: o próprio autor do texto.

Especificamente para a cidade de Jacarta, a qual é reportada com elevado percentual de uso de motocicletas (SUSILO et al., 2007) foi utilizado um consumo unitário diferenciado para o modo carro. Este consumo unitário diferenciado foi composto por 68.5% do consumo médio unitário para o modo carro comum para todas as cidades, conforme a Tabela 5.5, e por 31.5% de um modo moto com consumo unitário de 1,37 MJ/km. O consumo unitário do modo moto foi estimado com base no rendimento de 25 km/l de combustível gasolina, tendo em vista não se ter obtido uma estimativa da frota veicular de motocicletas naquela cidade.

O indicador de consumo de viagem devido à mobilidade urbana observada em uma cidade é o consumo médio de todas as viagens coletadas, calculado por média simples dos consumos de todas as viagens da amostra de uma cidade.

5.4.1 Resultados – Consumo Médio de Energia Devido à Mobilidade Urbana Observada em cada Cidade

O gráfico a seguir apresenta os indicadores de consumo de energia para modos motorizados devidos à mobilidade urbana observada em cada cidade enfocada neste estudo. Vale destacar que para o cálculo desses consumos médios para as cidades de São Paulo e Londres, embora se dispusesse das pesquisas origem-destino destas cidades, foram utilizadas as amostras coletas através de redes sociais, assim como foram para as demais, para fins comparativos.

Figura 5.14 Indicadores de consumo de energia devido à mobilidade urbana observada em cada cidade

Fonte: o próprio autor do texto.

Figura 5.15 Quantidade de viagens da amostra de cada cidade após filtragem para dados nulos e faltantes, e quantidade de viagens por modos motorizados utilizadas no cálculo dos consumos médios de energia devido à mobilidade urbana observada em cada cidade

Fonte: o próprio autor do texto.

Com o intuito de considerar a influência do modo caminhada na divisão modal e, portanto, no consumo de energia e consequentemente no ranque de consumo devido à mobilidade urbana, foram calculados os indicadores de consumo de cada cidade também considerando o número de viagens por modo caminhada. A Figura 5.16 apresenta os

indicadores do consumo de energia devido à mobilidade urbana calculados para todos os modos de viagem.

Figura 5.16 Indicadores de energia devido à mobilidade urbana por todos os modos de viagem, incluído o modo caminhada, observada em cada cidade

Fonte: o próprio autor do texto.

No Apêndice C encontram-se histogramas de distâncias de viagens por modos motorizados de São Paulo e Londres, tanto das pesquisas origem-destino como das viagens provenientes das redes sociais, para uma observação comparativa entre essas duas diferentes fontes de dados. Nota-se que há similaridade entre as curvas dos dados dessas duas fontes para um mesmo modo de viagem, o que exprime a boa performance preditiva do modelo RF ajustado. Ou seja, a análise dos histogramas permite inferir que a atribuição dos modos de viagem realizada pelo modelo de escolha ajustado, o algoritmo RF treinado, para as viagens oriundas da base de dados das redes sociais, se aproxima do observado nos dados das pesquisas origem-destino, para aquelas duas cidades, Londres e São Paulo.

As figuras a seguir apresentam as divisões componentes do cálculo do consumo médio de energia devido à mobilidade urbana em cada cidade, bem como a divisão modal de modos motorizados, resultado da predição de um modo de viagem a cada uma das viagens, de cada cidade, apresentadas ao modelo classificador *Random Forest* anteriormente escolhido e treinado, conforme linhas acima neste texto.

A respeito da cidade de Nova Iorque, as figuras apresentam a maior participação percentual do modo metrô no cálculo do indicador de consumo médio dentre todas as cidades.

Figura 5.17 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Nova Iorque

Fonte: o próprio autor do texto.

Figura 5.18 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Nova Iorque

Fonte: o próprio autor do texto.

A divisão modal estimada para Nova Iorque pelo modelo RF ajustado, Figura 5.19, apresenta as maiores quotas percentuais para os modos metrô e caminhada dentre todas as oito cidades enfocadas, à semelhança do apresentado na Tabela 5.2 acima, conforme fontes descrevem.

Figura 5.19 Divisão modal estimada pelo modelo de escolha algoritmo RF para Nova Iorque

Fonte: o próprio autor do texto.

Embora a cidade de Paris figure em segundo lugar no ranque dos indicadores de consumo de energia devido à mobilidade urbana da Figura 5.14, as participações percentuais dos modos carro e metrô são bem diferente daquelas estimadas para a cidade com o menor indicador de consumo, Nova Iorque.

Figura 5.20 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Paris

Fonte: o próprio autor do texto.

Figura 5.21 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Paris

Fonte: o próprio autor do texto.

Figura 5.22 Divisão modal estimada pelo modelo de escolha algoritmo RF para Paris

Fonte: o próprio autor do texto.

A cidade do Rio de Janeiro apresenta a maior participação percentual do modo ônibus para o cálculo do indicador de consumo de energia devido à mobilidade urbana dentre as cidades, sendo também a maior quota percentual do modo ônibus na divisão modal estimada pelo modelo RF ajustado, o que concorda com o apresentado na Tabela 5.2. Tal grande participação do modo ônibus na mobilidade observada no Rio de Janeiro promove sua terceira colocação do ranque da Figura 5.14.

Figura 5.23 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Rio de Janeiro

Fonte: o próprio autor do texto.

Figura 5.24 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Rio de Janeiro

Fonte: o próprio autor do texto.

Figura 5.25 Divisão modal estimada pelo modelo de escolha algoritmo RF para o Rio de Janeiro

Fonte: o próprio autor do texto.

É expressiva a diferença de participação percentual do modo carro no cálculo do indicador de consumo de energia da quarta cidade colocada no ranque da Figura 5.14, Toronto, e suas vizinhas naquele ranque, Rio de Janeiro e Jacarta.

Figura 5.26 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Toronto

Fonte: o próprio autor do texto.

Figura 5.27 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Toronto

Fonte: o próprio autor do texto.

Figura 5.28 Divisão modal estimada pelo modelo de escolha algoritmo RF para Toronto

Fonte: o próprio autor do texto.

Para a Figura 5.29, a seguir, vale lembrar que o modo carro para a cidade de Jacarta é misto, composto por carro e motocicleta, o que atenua a participação de tal modo carro na divisão do consumo de energia devido à mobilidade urbana.

Figura 5.29 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Jacarta

Fonte: o próprio autor do texto.

Figura 5.30 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Jakarta

Fonte: o próprio autor do texto.

Figura 5.31 Divisão modal estimada pelo modelo de escolha algoritmo RF para Jakarta

Fonte: o próprio autor do texto.

Para os componentes dos indicadores de consumo de energia de Londres e São Paulo, a participação percentual do modo ônibus em Londres é aproximadamente a metade daquela calculada para São Paulo. Enquanto que a participação percentual do modo metrô em Londres é aproximadamente o dobro daquela calculada para São Paulo – vide Figuras 5.32 e 5.35.

Figura 5.32 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Londres

Fonte: o próprio autor do texto.

Figura 5.33 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Londres

Fonte: o próprio autor do texto.

Figura 5.34 Divisão modal estimada pelo modelo de escolha algoritmo RF para Londres

Fonte: o próprio autor do texto.

Figura 5.35 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de São Paulo

Fonte: o próprio autor do texto.

Figura 5.36 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de São Paulo

Fonte: o próprio autor do texto.

Figura 5.37 Divisão modal estimada pelo modelo de escolha algoritmo RF para São Paulo

Fonte: o próprio autor do texto.

O indicador de consumo de energia devido à mobilidade urbana calculado para a cidade de Los Angeles se deve quase que exclusivamente ao modo carro para a mobilidade urbana observada, como mostra a Figura 5.38.

Figura 5.38 Divisão componentes do indicador de consumo de energia devido à mobilidade urbana de Los Angeles

Fonte: o próprio autor do texto.

Figura 5.39 Divisão modal de modos motorizados utilizados no cálculo do indicador de consumo de energia devido à mobilidade urbana de Los Angeles

Fonte: o próprio autor do texto.

Figura 5.40 Divisão modal estimada pelo modelo de escolha algoritmo RF para Los Angeles

Fonte: o próprio autor do texto.

6 ANÁLISE ESPACIAL

O capítulo anterior apresentou o ranque das oito cidades enfocadas neste estudo, conforme os respectivos consumos de energia devido às suas mobilidades observadas, sendo a de menor consumo a cidade de Nova Iorque e a de maior consumo a cidade de Los Angeles. Neste capítulo é conduzida uma análise espacial exploratória para o atendimento de um dos objetivos deste estudo, que é a busca por diretrizes sustentáveis observadas na cidade de menor consumo energético.

6.1 Considerações Iniciais

Na literatura consultada a respeito da relação entre uso de energia e mobilidade urbana, a maioria dos estudos relaciona o consumo de energia à mobilidade urbana e à densidade urbana onde aquela se realiza, NEWMAN & KENWORTHY (1989a, 1989b) MIDALI et al. (2004), BANISTER et al. (1997), SUSILO & STEAD (2008), LIU & SHEN (2011), LE NÉCHET (2012). Tais abordagens lidam com o aspecto morfológico apenas da cidade, sua área, extensão e forma, para uma relação com uma grandeza física derivada, a energia consumida para realizar um deslocamento no espaço urbano, a qual advém de mecanismos associados ao sistema complexo de fluxos de uma cidade. Nesse sentido, BREHENY (1992) apresenta crítica ao caráter superficial de tal relação entre a densidade média urbana e o consumo de energia devido a transportes.

Com efeito, nesse contexto está associado o conceito de cidade compacta como uma solução ao consumo de energia devido à mobilidade urbana que, para alguns apresenta-se como hipótese aceitável, ECOTEC (1993) apud BREHENY (1995) e NEWMAN & KENWORTHY (1992), e para outros não, como NEUMAN (2005, p.23) em “A Falácia da Cidade Compacta”, na qual descreve que “... a cidade compacta não é condição necessária ou suficiente para uma

cidade ser sustentável e que a tentativa de tornar as cidades mais sustentáveis somente usando estratégias de forma urbana é contraproducente.”.

Entre as situações limites de concentração de toda a população em um único ponto, onde não há deslocamentos, tampouco consumo de energia devido a esses, e a mesma população urbana esparramada em área proporcional a um Estado ou país, por hipótese, em alguma medida, o aumento da extensão do espaço urbano, mantendo-se a população constante – variação de densidade –, poderia estar relacionado a um aumento no consumo de energia devido à mobilidade urbana, pois sendo a mobilidade por modo caminhada limitada em termos de velocidade e alcance, os modos alternativos motorizados consomem proporcionalmente à distância percorrida.

A partir de outra perspectiva, NOULAS et al. (2012), através de evidência empírica, encontraram um padrão universal, em 34 cidades mundiais, de mobilidade humana baseado em medida da quantidade de locais (pontos de interesse ou *venues* como chamados na plataforma *Foursquare*) próximos em termos de distância ao ponto de origem, de forma que a probabilidade de haver um deslocamento para um ponto de destino, um local, diminui quando o número de locais mais próximos à origem que a um potencial destino aumenta. Tal relação encontrada se aproxima da ideia apresentada por Samuel Stouffer ao desenvolver sua teoria das oportunidades intervenientes na cidade de Cleveland, STOUFFER (1940). Conforme mencionado anteriormente neste texto, aqueles autores, NOULAS et al. (2012), apresentam que as variações de deslocamentos humanos nas cidades são principalmente devidas à distribuição espacial dos locais, pontos de interesse em uma cidade onde nossas atividades cotidianas ocorrem, em vez de outros potenciais fatores, tais como sócios-culturais, transportes, ou relacionados à cognição humana. Assim, haveria uma relação entre a mobilidade humana e a densidade de locais e seu arranjo no espaço urbano.

Essa noção de mobilidade nas cidades como um sistema complexo, no qual a interação de indivíduos com locais ou arranjo espacial das atividades urbanas é evidenciada como um padrão em larga escala mundial, está consistente com o novo paradigma de análise das cidades como sistemas complexos apresentado por BATTY(2008), em que a interação entre os elementos do sistema é tema central, emergindo-se padrões a partir de interações de forma de-baixo-para-cima, *bottom up*.

Portanto, em busca do atendimento a um dos objetivos deste estudo é apresentada a seguir uma investigação sobre a relação do consumo de energia da mobilidade observada por modos motorizados em algumas cidades e a densidade de locais nelas, em busca de informação útil. Foram enfocadas duas cidades, a de menor consumo, Nova Iorque, e a de maior consumo Los Angeles, na direção de que uma busca comparativa pudesse trazer informações úteis como as diretrizes sustentáveis almejadas neste estudo. Essa investigação será realizada sob a ótica da análise espacial tendo em vista a provável relação entre o fenômeno a ser observado, o consumo de energia devido ao deslocamento no espaço urbano e as feições do próprio espaço urbano que com ele interagem.

6.2 Análise Espacial

6.2.1 Análise espacial exploratória

A Análise Espacial, ou Estatística Espacial, se configura por um conjunto de técnicas ou métodos quantitativos aplicados a conjunto de dados, nos quais cada instância do conjunto tem uma referência geográfica, ou seja, sabe-se onde cada instância ocorre, daí o nome espacial. A diferença da análise ou estatística comum é a inclusão da interação do fenômeno focado e o espaço onde ele se realiza (FISCHER & GETIS, 2010).

A análise espacial a ser aplicada é conhecida como Análise Exploratória de Dados Espaciais (ESDA, em inglês), também considerada uma análise dirigida a dados (ANSELIN & REY, 2009), e é destinada à descoberta de padrões espaciais para a sugestão de hipóteses, onde a apresentação de padrões ou agrupamentos de dados, *clusters* é acompanhada de nível de significância estatístico, o que a diferencia da análise visual humana de padrões, a qual é sujeita a falhas.

Para a análise pretendida entre a energia gasta na realização de viagens motorizadas e a densidade de locais, faz-se necessária a escolha da malha, *grid*, a ser computada tal densidade. A escolha do tamanho da malha é realizada com base na variância espacial ou autocorrelação espacial do fenômeno a ser estudado, e essa seguiu um método apresentado em FORTIN & DALE (2005). O tamanho da malha, ou dos lados de suas células (como por exemplo, uma malha quadrada de células com 500 metros por 500 metros de medida de seus lados) interfere nos valores da variância espacial, e é uma questão conhecida na área de análise espacial como “Problema da Unidade de Área Modificável” (MAUP, em inglês, OPENSHAW, 1984).

Portanto, utiliza-se um método de escolha da malha em face da variância espacial, do atributo ou atributos a ser(em) estudado(s) determinando, então, qual o tamanho da malha a ser utilizado.

A variância espacial, ou padronagem espacial, ou como é mais frequente chamada, autocorrelação espacial, de um atributo, pode ser expressa como a coincidência de proximidade em localização com similaridade de valor desse atributo. Essa propriedade é quantificada por uma medida estatística utilizada para verificar a hipótese nula de que os dados estão aleatoriamente distribuídos no espaço. Existem diferentes índices ou fórmulas para estimar a autocorrelação espacial. Este estudo empregará, talvez, a mais conhecida delas, denominada estatística Moran's I, ou Global Moran's I, (MORAN, 1948), que tem como hipótese subjacente a completamente aleatória distribuição espacial do fenômeno ou atributo em teste, e pode ser expressa como: o valor de um atributo em um lugar não depende dos valores de um atributo na vizinhança desse lugar – no caso da autocorrelação espacial, o mesmo atributo.

O índice estatístico Moran's I, que se formaliza na Equação 6.1 a seguir, varia no intervalo $[-1,1]$, sendo que um valor próximo de -1 indica a presença de forte autocorrelação espacial negativa entre os valores do atributo em análise (caso do tabuleiro do jogo de xadrez), e um valor próximo de 1 indica forte autocorrelação positiva entre os valores do atributo e agrupamento, *clustering*. Por sua vez, valores próximos de zero indicam aleatoriedade da distribuição espacial dos valores do atributo em análise.

$$I = \frac{N}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad i \neq j \quad (\text{Eq. 6.1})$$

Onde n é o número de locais, x_i é o valor do atributo em análise, W_{ij} é um peso espacial para o par de locais i e j . A definição da matriz de pesos espaciais depende da consideração sobre de qual forma se dá a vizinhança de um local. Existem diferentes convenções para a vizinhança de um local, como exemplo a chamada “rainha” que, em alusão ao movimento daquela peça do jogo de xadrez, considera os oito vizinhos de uma célula de uma malha quadrada. Ou a chamada “torre”, que também correspondendo ao movimento em cruz daquela peça no jogo de xadrez, considera os quatro vizinhos de uma célula quadrada em uma malha. Essas convenções e a estatística Moran's I se baseiam na lei da Geografia conhecida como a lei de Tobler (TOBLER,

1970), a qual enuncia que tudo está relacionado com tudo, mas as coisas próximas estão mais relacionadas do que as distantes.

O método utilizado para a escolha do tamanho da malha observa em qual das malhas, já tendo sido inseridos os pontos com seus atributos dentro das células, obtém-se o maior valor da estatística Moran's I, ou seja, em qual tamanho de malha se observa a maior correlação espacial. Em havendo dois pontos de pico, como exemplo para as malhas 200x200 e 600x600, escolhe-se, no caso deste estudo e em geral, a malha de menor tamanho em favor de uma maior variância do atributo em sua distribuição espacial, uma vez que a agregação em maior dimensão caminha no sentido de redução de informação (HENGL, 2006).

Porém, antes de inserir os dados das cidades nas diferentes malhas, há que ser considerada uma questão sobre a cidade de Nova Iorque. A peculiar distribuição geográfica de forma insular associada à também peculiares feições urbanas, como a do sistema de transportes, viário e suas pontes de ligação, a de ocupação imobiliária, etc, caracteriza violação à propriedade de estacionariedade (FORTIN & DALE, 2005) necessária à investigação por análise espacial com inferência estatística aqui intencionada. Exemplificando, assim como não se pode ajustar um modelo espacial de precipitação pluviométrica para ambos os lados de uma região com uma íngreme cordilheira de montanhas ao meio, pois a assunção de que o processo a ser modelado é estacionário, homogêneo em ambos os lados, não se aplica por implicação de que o padrão espacial encontrado e o modelo ajustado podem estar distorcidos e imprecisos.

A observação da Figura 6.1 indica que o bairro chamado Ilha Staten, bem como o território à margem esquerda da ilha de Manhattan, não apresentam o mesmo padrão de sistema de transportes, além de haver a interrupção da continuidade territorial em grande medida e com poucas pontes de acesso para interligação com o bloco principal constituído pelos bairros Manhattan, Bronx, Brooklyn e Queens.

Figura 6.1 Vista ampliada dos bairros de Nova Iorque com seu sistema de transportes

Fonte: o próprio autor do texto.

Também é possível notar na Figuras 6.2 que, tanto na Ilha Staten como no território à margem esquerda de Manhattan, há uma a quantidade menor de locais de destino de viagens, sendo que o local de destino de uma viagem é a localização para o atributo de energia consumida na viagem na análise espacial conduzida neste estudo.

Portanto, apesar de haver descontinuidade territorial entre eles e diferentes tipos de ocupação territorial, dada a singularidade de Manhattan, para fins da análise espacial com inferência estatística, a seguir apresentada, foram considerados apenas os bairros de Nova Iorque mais próximos entre si e com maior semelhança em sua feição geográfica de transportes, sendo eles, Manhattan, Bronx, Brooklyn e Queens, ainda que houvesse a perda de algumas instâncias de viagens.

Figura 6.2 Vista ampliada dos bairros de Nova Iorque com locais de destino de viagens

Fonte: o próprio autor do texto.

A respeito da descontinuidade territorial da cidade de Nova Iorque, a única malha testada à estatística Moran's I foi a de tamanho de célula de 200x200 metros, pois a partir do tamanho de malha 400x400 metros surgem continuidades e vizinhanças onde na verdade não existem tais. As Figuras 6.3, 6.4 e 6.5 a seguir ilustram essa situação.

Figura 6.3 Vista em detalhe da malha 200x200 metros sobre sistema de transportes de Nova Iorque

Fonte: o próprio autor do texto.

Figura 6.4 Vista em detalhe da malha 400x400 metros sobre sistema de transportes de Nova Iorque. Essa malha 400x400 metros já apresenta continuidade e vizinhanças de células além do real territorial.

Fonte: o próprio autor do texto.

Figura 6.5 Vista em detalhe da malha 600x600 metros sobre sistema de transportes de Nova Iorque

Fonte: o próprio autor do texto.

Para o cálculo da estatística Moran's I foram considerados 7 tipos de vizinhança, são eles: a rainha; a rainha em segunda ordem, ou seja, com mais uma coroa de vizinhos além dos oito mais próximos; a torre; duas vizinhanças baseadas na distância de raios de 2.000 e 4.000 metros, tomados o valores das distâncias de forma inversa; e duas vizinhanças do tipo vizinhos mais próximos, uma com 4 e outra com 8 vizinhos.

Figura 6.6 Gráfico dos valores da estatística Moran's I calculados para a malha de células quadradas, de medidas 200 por 200 metros, e para a média de energia consumida nas viagens inseridas em cada célula

Fonte: o próprio autor do texto.

Figura 6.7 Gráfico dos valores da estatística Moran's I calculados por malha para a média de energia consumida nas viagens inseridas em cada célula (significância pseudo valor-p =0.001)

Fonte: o próprio autor do texto.

A partir da análise dos gráficos das Figuras 6.6 e 6.7, foi escolhida a malha de 200x200 metros para ambas as cidades, tendo em vista que, embora os valores do índice Moran's I apresentem pico para a malha 600x600 metros, devido à possibilidade de perda de informação por agregação em ordem superior. Também a partir da análise das mesmas figuras, foi escolhida a vizinhança torre por apresentar valores mais expressivos de Moran's I e gerar matriz de pesos

simétrica, possibilitando maior agilidade no processamento computacional, diferentemente da vizinhança 4NN.

Também a partir das Figuras 6.6 e 6.7, pode-se notar que há autocorrelação, ou padronagem, espacial para o atributo testado energia média consumida em viagens com destino contido nas células de forma evidenciada em Los Angeles, e menor em Nova Iorque. Porém, apesar da estatística Moran's I calculada para Nova Iorque apresentar recusa à hipótese nula de distribuição aleatória do atributo energia média, consumida em viagens com destinos contidos na célula, vide Figura 6.8, chama a atenção os valores baixos dessa estatística.

Tal fato pode ser atribuído parcialmente à característica geográfica insular com descontinuidade territorial daquela cidade, pois após tal constatação de baixos valores, foram testadas malhas de dimensões maiores e gradualmente ao se aumentar a malha, e por conseguinte eliminação gradual da descontinuidade territorial, observou-se elevação correspondentemente gradual da estatística Moran's I, atingindo valor superior a 0.15 na malha de 800x800 metros. Para apenas a ilha de Manhattan, sem o bairro próximo Bronx, o valor da estatística Moran's I calculado foi de 0.09. Ainda, para o atributo energia média, o padrão observado de separação clara entre um principal agrupamento e demais pequenos agrupamentos escassos e esparramados na área territorial enfocada, contribui para o valor baixo obtido da estatística Moran's I de Nova Iorque.

Figura 6.8 Teste de significância da estatística Moran's I para o atributo energia media consumida por célula da malha de Nova Iorque

Fonte: o próprio autor do texto.

As figuras a seguir apresentam o mapa das duas cidades com os dados dos atributos energia consumida nas viagens, ambos já inseridos na malha 200 escolhida. A Tabela 6.1 apresenta um sumário de estatística descritiva para esses atributos.

Figura 6.9 Mapa de Nova Iorque com o atributo energia consumida nas viagens com destino contido em célula da malha 200 por 200 metros

Fonte: o próprio autor do texto.

Figura 6.10 Mapa de Los Angeles com o atributo quantidade de locais por célula da malha 200 por 200 metros

Legenda

Los Angeles malha 200

Energia média consumida viagens destinos na célula [MJ]

Fonte: o próprio autor do texto.

Tabela 6.1 Sumário de estatística descritiva* dos atributos energia consumida nas viagens com destino na célula e quantidade de locais por célula

		Energia média consumida viagem, destinos na célula	Quantidade de locais por célula
Nova Iorque	média	1994.3*	5.5
	desv. Padrão	5756.9	16.1
	mínimo	0.0	0
	máximo	194380.0	268
	mediana	0.0	1
	assimetria (skewness)	8.5	6.6
	kurtosis	148.3	58.8
Los Angeles	média	6101.9*	1
	desv. Padrão	19666.0	3.51
	mínimo	0.0	0
	máximo	296850.0	98
	mediana	0.0	0
	assimetria (skewness)	4.4	8.7
	kurtosis	27.2	124.3

Fonte: o próprio autor do texto. * Medidas estatísticas calculadas para toda a malha, portanto entrando no cálculo as células vazias com valores zero para os atributos desta tabela.

6.2.2 Análise de agrupamentos

Para a identificação de agrupamentos espaciais (*spatial clusters*) será utilizado um indicador local de associação espacial, proposto por ANSELIN (1995), o qual, de forma semelhante a estatística Moran's I, testa com base na hipótese de nulidade de que é aleatória a distribuição local do valor de um atributo para a associação em agrupamento. Tal indicador se chama Moran's I Local, ou LISA Moran's Local, e se formaliza na Equação 6.2, a seguir apresentada. Além da função de apresentar a localização de agrupamentos espaciais, esse indicador estatístico também apresenta os *outliers* espaciais, valores atípicos espaciais, os quais diferem dos *outliers* da estatística padrão, pois os espaciais são definidos como observações que não seguem o mesmo processo de dependência espacial da maioria do conjunto de dados,

e, no caso do local, da vizinhança ANSELIN (1996). Os outliers espaciais são identificados nos mapas com cores e legendas *High-Low* e *Low-High* (mantidas em inglês pois não há possibilidade de alteração desses textos das legendas dos mapas no software utilizado), para alto-baixo e baixo-alto, do idioma inglês do software utilizado.

$$I_i = \frac{z_i - \bar{z}}{\sigma^2} \sum_j^n [w_{ij}(z_j - \bar{z})] \quad (\text{Eq6.2})$$

Onde z_i é o valor do atributo z na localização i ; $\bar{z}$ é a média da amostra de n elementos; z_j é o valor do atributo z na vizinhança, $j \neq i$; σ^2 é a variância da variável z ; e w_{ij} é um peso definido como o inverso da distância entre as localizações i e j .

A figuras a seguir apresentam mapas resultantes da análise da agrupamentos para as cidades de Nova Iorque e Los Angeles e para os atributos em análise.

Figura 6.11 Mapa de agrupamentos LISA LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Nova Iorque (não significativo representado na cor cinza)

Fonte: o próprio autor do texto.

Figura 6.12 Mapa de índice de significância dos agrupamentos LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Nova Iorque

Fonte: o próprio autor do texto.

Figura 6.13 Gráfico do indicador LISA Moran's I para os agrupamentos do atributo energia média consumida em viagens com destino na célula, Nova Iorque

Fonte: o próprio autor do texto.

Figura 6.14 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Nova Iorque

- BiLISA Cluster Map
- Not Significant (13718)
 - High-High (642)
 - Low-Low (1102)
 - Low-High (90)
 - High-Low (163)

Fonte: o próprio autor do texto.

Figura 6.15 Mapa de índice de significância dos agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Nova Iorque

Fonte: o próprio autor do texto.

Figura 6.16 Gráfico do indicador LISA Moran's I para os agrupamentos do atributo quantidade de locais por célula, Nova Iorque

Fonte: o próprio autor do texto.

Figura 6.17 Mapa de agrupamentos LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Los Angeles

Fonte: o próprio autor do texto.

Figura 6.18 Mapa de índice de significância dos agrupamentos LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Los Angeles

Fonte: o próprio autor do texto.

Figura 6.19 Gráfico do indicador LISA Moran's I para os agrupamentos do atributo energia média consumida em viagens com destino na célula, Los Angeles

Fonte: o próprio autor do texto.

Figura 6.20 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Los Angeles

Fonte: o próprio autor do texto.

Figura 6.21 Mapa de índice de significância dos agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Los Angeles

Fonte: o próprio autor do texto.

Figura 6.22 Gráfico do indicador LISA Moran's I para os agrupamentos do atributo quantidade de locais por célula, Los Angeles

Fonte: o próprio autor do texto.

Figura 6.23 Mapa de agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Nova Iorque

Fonte: o próprio autor do texto.

Figura 6.24 Mapa de índice de significância dos agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Nova Iorque

Fonte: o próprio autor do texto.

Figura 6.25 Gráfico do indicador LISA Moran's I bivariado para os agrupamentos dos atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Nova Iorque

Fonte: o próprio autor do texto.

Figura 6.26 Mapa de agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Los Angeles

Fonte: o próprio autor do texto.

Figura 6.27 Mapa de índice de significância dos agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Los Angeles

Fonte: o próprio autor do texto.

Figura 6.28 Gráfico do indicador LISA Moran's I bivariado para os agrupamentos dos atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Los Angeles

Fonte: o próprio autor do texto.

Da análise dos mapas de agrupamentos acima para o atributo de energia consumida em viagens, percebe-se uma nítida diferença entre as duas cidades. Em Nova Iorque praticamente não há agrupamento de área considerável para o atributo de energia, sendo o maior agrupamento percebido na ponta sul da ilha de Manhattan. Por outro lado, em Los Angeles pode-se notar uma maior quantidade de agrupamentos, com áreas maiores que aqueles observados na outra cidade para o atributo energia. Ainda, as localizações dos agrupamentos do atributo energia correspondem em certa medida às localizações dos agrupamentos do atributo quantidade de locais. Tal fato é também observável no mapa bivariado de agrupamentos de Los Angeles, no qual a análise conjunta de ambos os atributos gera um mapa semelhante ao do atributo quantidade de locais e/ou do atributo energia, devido a semelhança entre eles. Porém, o mesmo não se observa no mapa bivariado de Nova Iorque, no qual a junção de ambos os atributos não se assemelha ao mapa do atributo quantidade de locais apenas.

Como resultado dessa análise dos mapas de agrupamentos para o atributo de energia, induz-se a ideia de que em havendo facilidade do transporte público de baixo consumo unitário, metrô, Transporte Rápido de Massa (MRT, em inglês), de forma abrangente como o observado em Nova Iorque, passa-se a não haver correspondência entre a densidade de locais e a energia média consumida para se chegar aos pontos de maior densidade de locais, uma vez que não há em Nova Iorque correspondência de agrupamentos para os atributos sob análise. Ao passo que em Los Angeles, com reduzida extensão da facilidade metrô, ocorre tal correspondência.

Apesar de a quantidade de locais que se dispõe neste estudo, 103.316 locais para Nova Iorque e 51.317 locais para Los Angeles, não ser a quantidade total de locais existentes em cada uma das cidades, mas apenas uma parcela dela, a partir da análise dos mapas de agrupamentos do atributo quantidade de locais, fica evidente a tendência monocêntrica em Nova Iorque e a policêntrica em Los Angeles. Os pequenos agrupamentos de locais dispersos em grande parte da área da cidade de Los Angeles se contrapõem à grande e quase única área de agrupamento de locais em Manhattan. Tal constatação é consistente com o apresentado por BERTAUD (2001, 2003) a respeito dessas duas cidades, Nova Iorque e Los Angeles. A relação apresentada na Figura 6.29 a seguir, extraída de LEFEVRE (2009), corresponde ao observado neste estudo a respeito do uso de transporte público na mobilidade urbana observada, a divisão modal aqui estimada, e da estrutura espacial, no que tange a variação entre monocentrismo e policentrismo, a dispersão de pequenos agrupamentos de locais obtida neste estudo, para a cidade de Los Angeles.

Figura 6.29 Relação entre estrutura espacial e a efetividade de transporte público

Fonte: BERTAUD & MALPEZZI (2003) apud LEFEVRE (2009).

Com vistas a melhor observar o fenômeno do agrupamento de locais, como um fenômeno *proxy* da distribuição espacial, mono ou policêntrica, as figuras a seguir apresentam os mapas de agrupamentos para as demais seis cidades enfocadas neste estudo, por ordem de consumo de energia conforme o rank obtido no capítulo anterior.

Figura 6.30 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Paris

Fonte: o próprio autor do texto.

Figura 6.31 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Rio de Janeiro

Fonte: o próprio autor do texto.

Figura 6.32 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Toronto

Fonte: o próprio autor do texto.

Figura 6.33 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Jacarta

Fonte: o próprio autor do texto.

Figura 6.34 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, Londres

Fonte: o próprio autor do texto.

Figura 6.35 Mapa de agrupamentos LISA Moran's I para atributo quantidade de locais por célula, São Paulo

Fonte: o próprio autor do texto.

Da análise dos mapas de agrupamentos dessas outras seis cidades, pode-se observar primeiramente que é nítida a situação monocêntrica de Paris, assim como de Nova Iorque, como se essas duas cidades, as mais eficientes, conforme este estudo, em termos de consumo de

energia devido à mobilidade urbana, formassem um grupo isolado das demais, o que também se observa em termos do indicador de consumo obtido no capítulo anterior.

A cidade do Rio de Janeiro apresenta padrão policêntrico de distribuição espacial ou centralidade, entretanto há de ser considerado o aspecto geográfico territorial característico daquela cidade litorânea e com acidentes geográficos, e que apresentou um valor baixo do indicador de consumo energético no ranque do capítulo anterior, devido principalmente à parcela de uso do modo de transporte ônibus.

A partir da cidade de Toronto observa-se, exceto em Jacarta, um misto de centralidade monocêntrica com algum grau de policentrismo, com ênfase ao policentrismo enunciado na cidade de São Paulo. Essa variação é consistente com o indicador de consumo energético obtido no capítulo anterior, para o caso de se relacionar tal consumo de energia devido à mobilidade urbana apenas com a distribuição espacial dos locais da cidade.

Jacarta apresenta-se como um caso distinto, não apresentando agrupamento central de tamanho considerável em relação a sua área urbana, tampouco apresentando agrupamentos consideráveis em localizações distantes a sua área central. De fato, o que se observa em Jacarta é misto entre o monocentrismo e o policentrismo, ambos enunciados em grau menor que as outras cidades, como se estivesse em formação. O valor do indicador de agrupamento LISA Moran's I Local reflete essa situação tênue, sendo 0,19 para Jacarta contra, por exemplo, 0,73 para Paris. Sua distinção dentre as demais cidades aqui estudadas segue um padrão observado por BARTER (1999) para algumas cidades asiáticas que adotaram o modo de transporte urbano motocicleta em larga escala. Esse fato foi considerado para o cálculo do indicador de consumo do capítulo anterior, e pode estar relacionado à forma dos agrupamentos apresentados no seu mapa LISA Moran's I para o atributo quantidade de locais.

6.2.3 Regressão Espacial

A modelagem de regressão espacial neste estudo seguiu o método contido em ANSELIN (2005) para a busca pelo modelo de melhor ajuste. Destaca-se que o caráter da modelagem espacial neste estudo não é o de busca extensiva por um melhor modelo ajustado, como o que foi realizado no capítulo anterior, pois não há a intenção de utilização do modelo para uso em previsão. Mas sim, apenas a observação da existência ou não de relação entre as

variáveis, ou atributos, relacionados no modelo, e em qual nível de determinação do modelo ajustado isso se dá.

Os modelos a serem investigados são: o modelo clássico linear (OLS, em inglês), e os modelos espaciais, o modelo Simultâneo Autoregressivo (SAR, em inglês) ou também conhecido como modelo *Lag* Espacial (SLM), e o modelo de Erro Espacial (SEM, em inglês).

Uma vez que haja autocorrelação espacial na variável dependente, e esta seja estimada por um modelo clássico linear, surge autocorrelação espacial observável nos resíduos do modelo clássico ajustado. A ideia implícita nos modelos espaciais é a de que a autocorrelação espacial seja explicada por um termo da equação desses modelos, eliminando a autocorrelação antes observada nos resíduos do modelo clássico linear. Mais detalhes podem ser encontrados em FISCHER & GETIS(2010), ANSELIN(1988, 2005). Os modelos a serem utilizados, contidos no método seguido, estão formalizados de forma simplificada matricial a seguir.

Modelo OLS:

$$Y = X\beta + \varepsilon \quad (\text{Eq. 6.2})$$

Onde Y é o vetor da variável ou atributo dependente, X é a matriz das variáveis independentes, β é o vetor de parâmetros a serem estimados, e ε se refere ao vetor de termos de erro normalmente distribuído, com média igual a zero e variância constante.

Modelo SLM:

$$Y = \rho WY + X\beta + \varepsilon \quad (\text{Eq. 6.3})$$

Onde W é a matriz de pesos espaciais, ρ é o coeficiente espacial autoregressivo a ser estimado.

Modelo SEM:

$$Y = X\beta + u \quad , \quad u = \lambda Wu + \varepsilon \quad (\text{Eq. 6.4})$$

Onde λ é o coeficiente autoregressivo de resíduos a ser ajustado.

O método seguido na análise conduzida parte da obtenção de um modelo OLS ajustado para os dados, que é considerado como a hipótese nula de que esse é melhor modelo para o fenômeno em modelagem. É, então, verificada a existência de autocorrelação nos resíduos do modelo, e se for o caso, realizada a verificação de quatro testes estatísticos de Multiplicador

Lagrange, e suas respectivas significâncias (valores p), sendo dois testes pertencentes ao modelo *Lag* Espacial como hipótese alternativa, e dois pertencentes ao modelo de Erro Espacial como hipótese alternativa. O fluxo do método encontra-se no Anexo A.

Para melhor fundamentar a investigação conduzida sobre a relação entre a energia consumida nas viagens e a densidade de locais, o escopo da modelagem por inferência estatística foi estendido a todas as oito cidades enfocadas neste estudo, tendo em vista o fato da estatística Moran's I para Nova Iorque ter apresentado baixo valor.

Tabela 6.2 Estatística Moran's I calculadas para o atributo energia média consumida em viagens com destinos inseridos dentro da célula de 200x200 metros, para as demais seis cidades enfocadas

Moran's I	
Paris	0.143
Rio de Janeiro	0.158
Toronto	0.138
Jakarta	0.190
Londres	0.117
São Paulo	0.175

Fonte: o próprio autor do texto.

Os resultados da tarefa de ajuste dos modelos estatísticos encontram-se nas tabelas a seguir. De forma a propiciar comparação entre as cidades, para todos os modelos espaciais ajustados foram utilizadas malhas de 200x200 metros e vizinhança tipo torre, conforme a análise exploratória, item 6.2.1, descrita acima neste texto. Devido aos baixos índices de determinação dos modelos obtidos quando foram ajustados às variáveis tomadas da forma direta, bem como às distribuições dessas variáveis e a presença de muitos elementos zero devido a células vazias, foram testadas algumas transformações, raiz cúbica e logaritmo, tanto na variável dependente, quanto na independente para uma aproximação à distribuição normal nessas distribuições, e um melhor coeficiente de determinação dos modelos.

Tabela 6.3 Resultados da regressão, cidade de Nova Iorque

		R ² ou pseudo- R ²	AIC	Jarque- Bera	Breusch- Pagan	Moran's I: resíduos da regressão
Nova Iorque	Mod. Variáveis					
	OLS Energia ~ Q. Locais	0.017	316483	1.5E+07	3	7.906
	OLS (Energia) ^{1/3} ~ Q. Locais	0.112	108660	11028.3	838	22.722
	OLS log(Energia) ~ Q. Locais	0.165	83135	3678	1946	29.026
	SEM Energia ~ Q. Locais	0.017	-	-	-	0.027
	SEM (Energia) ^{1/3} ~ Q. Locais	0.112	-	-	-	0.076
	SEM log(Energia) ~ Q. Locais	0.165	-	-	-	0.100
	OLS log(Energia) ~ log(Q. Locais)	0.335	79558	707	5363	11.662
	SLM log(Energia) ~ log(Q. Locais)	0.354	79231	-	5395	-0.033
	OLS log(Energia) ~ (Q. Locais) ^{1/3}	0.419	77450	1297	544	6.813
	SEM log(Energia) ~ (Q. Locais) ^{1/3}	0.419	-	-	-	0.025
	OLS s/ outliers log(Energia) ~ (Q. Locais) ^{1/3}	0.447	74205	1191	645	10.471
	SLM s/ outliers log(Energia) ~ (Q. Locais) ^{1/3}	0.454	74070	-	602	-0.001

Fonte: o próprio autor do texto.

Tabela 6.4 Resultados da regressão, cidade de Paris

		R ² ou pseudo- R ²	AIC	Jarque- Bera	Breusch- Pagan	Moran's I: resíduos da regressão
Paris	OLS Energia ~ Q. Locais	0.090	717299	1.3E+08	2728	17.781
	OLS (Energia) ^{1/3} ~ Q. Locais	0.284	206477	1416193	39348	63.279
	OLS log(Energia) ~ Q. Locais	0.327	148942	277111	58784	74.454
	SLM Energia ~ Q. Locais	0.102	716944	-	1927	-0.011
	SLM (Energia) ^{1/3} ~ Q. Locais	0.377	202736	-	14417	-0.038
	SLM log(Energia) ~ Q. Locais	0.438	143955	-	19806	-0.043
	OLS log(Energia) ~ log(Q. Locais)	0.580	131805	43105	100354	21.772
	SLM log(Energia) ~ log(Q. Locais)	0.608	129796	-	97649	-0.060
	OLS log(Energia) ~ (Q. Locais) ^{1/3}	0.631	127154	78916	20737	12.535
	SLM log(Energia) ~ (Q. Locais) ^{1/3}	0.636	126777	-	20477	-0.011
	OLS s/ outliers log(Energia) ~ (Q. Locais) ^{1/3}	0.688	114200	114150	23879	39.708
	SLM s/ outliers log(Energia) ~ (Q. Locais) ^{1/3}	0.743	108688	-	19400	-0.091

Fonte: o próprio autor do texto.

Tabela 6.5 Resultados da regressão, cidade do Rio de Janeiro

		R ² ou pseudo- R ²	AIC	Jarque- Bera	Breusch- Pagan	Moran's I: resíduos da regressão
Rio de Janeiro	OLS Energia ~ Q. Locais	0.061	620388	1.5E+07	4076	28.498
	OLS (Energia) ^{1/3} ~ Q. Locais	0.167	195863	336811	37581	56.196
	OLS log(Energia) ~ Q. Locais	0.205	137797	148037	60803	66.211
	SLM Energia ~ Q. Locais	0.095	619617	-	2605	-0.143
	SLM (Energia) ^{1/3} ~ Q. Locais	0.275	192836	-	16377	-0.039
	SLM log(Energia) ~ Q. Locais	0.344	133550	-	23346	-0.049
	OLS log(Energia) ~ log(Q. Locais)	0.623	115770	39894	75996	9.875
	SLM log(Energia) ~ log(Q. Locais)	0.631	115286	-	75040	-0.026
	OLS log(Energia) ~ (Q. Locais) ^{1/3}	0.633	114982	45307	26231	13.074
	SEM log(Energia) ~ (Q. Locais) ^{1/3}	0.633	-	-	-	0.054
	OLS s/ outliers log(Energia) ~ (Q. Locais) ^{1/3}	0.679	104296	69898	34240	30.112
	SLM s/ outliers log(Energia) ~ (Q. Locais) ^{1/3}	0.709	102067	-	27436	-0.031

Fonte: o próprio autor do texto.

Tabela 6.6 Resultados da regressão, cidade de Toronto

		R ² ou pseudo- R ²	AIC	Jarque- Bera	Breusch- Pagan	Moran's I: resíduos da regressão
Toronto	OLS Energia ~ Q. Locais	0.055	332621	479399	1324	20.326
	OLS (Energia) ^{1/3} ~ Q. Locais	0.144	111391	38145	7677	33.134
	OLS log(Energia) ~ Q. Locais	0.199	78284	17479	13472	39.378
	SLM Energia ~ Q. Locais	0.085	332278	-	909	-0.003
	SLM (Energia) ^{1/3} ~ Q. Locais	0.214	110463	-	4318	-0.010
	SLM log(Energia) ~ Q. Locais	0.291	76922	-	6825	-0.016
	OLS log(Energia) ~ log(Q. Locais)	0.645	65938	10377	24450	2.577
	SLM log(Energia) ~ log(Q. Locais)	0.647	65858	-	24316	-0.021
	OLS log(Energia) ~ (Q. Locais) ^{1/3}	0.630	66548	6183	8876	11.931
	SEM log(Energia) ~ (Q. Locais) ^{1/3}	0.630	-	-	-	0.070
	OLS s/ outliers log(Energia) ~ log(Q. Locais)	0.686	59424	17752	25401	6.855
	SLM s/ outliers log(Energia) ~ log(Q. Locais)	0.707	58631	-	24322	-0.080

Fonte: o próprio autor do texto.

Tabela 6.7 Resultados da regressão, cidade de Jakarta

			R ² ou pseudo- R ²	AIC	Jarque- Bera	Breusch- Pagan	Moran's I: resíduos da regressão
Jakarta	OLS	Energia ~ Q. Locais	0.015	473021	271877	971	36.722
	OLS	(Energia) ^{1/3} ~ Q. Locais	0.033	167698	6279	4642	62.235
	OLS	log(Energia) ~ Q. Locais	0.040	122860	3267	6539	70.336
	SLM	Energia ~ Q. Locais	0.093	471775	-	562	-0.023
	SLM	(Energia) ^{1/3} ~ Q. Locais	0.222	164275	-	2035	-0.056
	SLM	log(Energia) ~ Q. Locais	0.268	118517		2657	-0.068
	OLS	log(Energia) ~ log(Q. Locais)	0.156	120099	5804	309	54.952
	SLM	log(Energia) ~ log(Q. Locais)	0.322	116555	-	178	-0.078
	OLS	log(Energia) ~ (Q. Locais) ^{1/3}	0.137	120587	5302	11	57.673
	SLM	log(Energia) ~ (Q. Locais) ^{1/3}	0.311	116943	-	4	-0.074
	OLS s/ outliers	log(Energia) ~ log(Q. Locais)	0.228	90175	29398	7	43.089
	SLM s/ outliers	log(Energia) ~ log(Q. Locais)	0.447	85657	-	4	-0.185

Fonte: o próprio autor do texto.

Tabela 6.8 Resultados da regressão, cidade de Londres

			R ² ou pseudo- R ²	AIC	Jarque- Bera	Breusch- Pagan	Moran's I: resíduos da regressão
Londres	OLS	Energia ~ Q. Locais	0.059	1128710	5.2E+07	14181	28.496
	OLS	(Energia) ^{1/3} ~ Q. Locais	0.152	363075	5.4E+05	89757	57.322
	OLS	log(Energia) ~ Q. Locais	0.195	246618	3.6E+05	141973	67.522
	SLM	Energia ~ Q. Locais	0.076	1128050	-	10329	0.001
	SLM	(Energia) ^{1/3} ~ Q. Locais	0.208	360532	-	49813	-0.005
	SLM	log(Energia) ~ Q. Locais	0.267	243094	-	71743	-0.009
	OLS	log(Energia) ~ log(Q. Locais)	0.513	220734	19704	112476	10.271
	SLM	log(Energia) ~ log(Q. Locais)	0.521	220050	-	111617	-0.034
	OLS	log(Energia) ~ (Q. Locais) ^{1/3}	0.548	216801	47830	32747	8.045
	SLM	log(Energia) ~ (Q. Locais) ^{1/3}	0.549	216768	-	32812	0.011
	OLS s/ outliers	log(Energia) ~ (Q. Locais) ^{1/3}	0.601	195331	67915	38462	33.483
	SLM s/ outliers	log(Energia) ~ (Q. Locais) ^{1/3}	0.634	192008	-	35032	-0.050

Fonte: o próprio autor do texto.

Tabela 6.9 Resultados da regressão, cidade de São Paulo

		R ² ou pseudo- R ²	AIC	Jarque- Bera	Breusch- Pagan	Moran's I: resíduos da regressão
São Paulo	OLS Energia ~ Q. Locais	0.087	784631	1926158	11444	29.998
	OLS (Energia) ^{1/3} ~ Q. Locais	0.258	166953	855661	121297	38.306
	OLS log(Energia) ~ Q. Locais	0.228	184800	106862	102282	64.622
	SLM Energia ~ Q. Locais	0.121	783697	-	6762	-0.018
	SLM (Energia) ^{1/3} ~ Q. Locais	0.316	164791	-	58582	-0.051
	SLM log(Energia) ~ Q. Locais	0.349	180203	-	36070	-0.061
	OLS log(Energia) ~ log(Q. Locais)	0.506	168912	1232	39247	16.809
	SLM log(Energia) ~ log(Q. Locais)	0.523	167918	-	38751	-0.040
	OLS log(Energia) ~ (Q. Locais) ^{1/3}	0.566	164313	5171	9019	7.863
	SLM log(Energia) ~ (Q. Locais) ^{1/3}	0.568	164170	-	8792	-0.007
	OLS s/ outliers log(Energia) ~ (Q. Locais) ^{1/3}	0.619	151853	7049	10635	23.797
	SLM s/ outliers log(Energia) ~ (Q. Locais) ^{1/3}	0.647	149867	-	8870	-0.058

Fonte: o próprio autor do texto.

Tabela 6.10 Resultados da regressão, cidade de Los Angeles

		R ² ou pseudo- R ²	AIC	Jarque- Bera	Breusch- Pagan	Moran's I: resíduos da regressão
Los Angeles	OLS Energia ~ Q. Locais	0.191	1115730	2586937	42251.9	29.416
	OLS (Energia) ^{1/3} ~ Q. Locais	0.297	365752	378435	137617	43.918
	OLS log(Energia) ~ Q. Locais	0.322	246450	302906	179343	48.549
	SLM Energia ~ Q. Locais	0.214	1114710	-	30954	-0.012
	SLM (Energia) ^{1/3} ~ Q. Locais	0.339	363538	-	91150	-0.015
	SLM log(Energia) ~ Q. Locais	0.370	243777	-	115839	-0.015
	OLS log(Energia) ~ log(Q. Locais)	0.489	232427	3036	69895	18.695
	SLM log(Energia) ~ log(Q. Locais)	0.506	231120	-	68745	-0.035
	OLS log(Energia) ~ (Q. Locais) ^{1/3}	0.562	224720	11858	18443	8.444
	SLM log(Energia) ~ (Q. Locais) ^{1/3}	0.565	224465	-	18466	-0.012
	OLS s/ outliers log(Energia) ~ (Q. Locais) ^{1/3}	0.616	205493	15941	17371	19.001
	SLM s/ outliers log(Energia) ~ (Q. Locais) ^{1/3}	0.644	202662	-	17300	-0.072

Fonte: o próprio autor do texto.

Tabela 6.11 Modelos clássicos OLS e espaciais SLM obtidos para a relação dos atributos conforme as tabelas anteriores, e após a retirada de outliers espaciais, considerando as relações dos atributos que apresentaram o maior coeficiente de determinação

Nova Iorque	OLS	$Y = X \ 2.377 + 0.156$
	SLM	$Y = 0.108 \ WY + X \ 2.202 - 0.221$
Paris	OLS	$Y = X \ 3.327 + 0.006$
	SLM	$Y = 0.368 \ WY + X \ 2.282 - 0.036$
Rio de Janeiro	OLS	$Y = X \ 3.876 + 0.050$
	SLM	$Y = 0.274 \ WY + X \ 3.054 - 0.011$
Toronto	OLS	$Y = X \ 0.867 + 5.951$
	SLM	$Y = 0.219 \ WY + X \ 0.745 + 4.999$
Jakarta	OLS	$Y = X \ 0.766 + 6.378$
	SLM	$Y = 0.501 \ WY + X \ 0.415 + 3.45$
Londres	OLS	$Y = X \ 3.860 + 0.002$
	SLM	$Y = 0.260 \ WY + X \ 3.134 - 0.054$
São Paulo	OLS	$Y = X \ 3.729 - 0.019$
	SLM	$Y = 0.241 \ WY + X \ 3.030 - 0.100$
Los Angeles	OLS	$Y = X \ 4.150 - 0.085$
	SLM	$Y = 0.244 \ WY + X \ 3.475 - 0.171$

Fonte: o próprio autor do texto.

Analisando-se as tabelas de resultados das regressões realizadas, exceto Nova Iorque e Jakarta, em todas as cidades obteve-se relações entre os atributos sob estudo com coeficiente de determinação do modelo acima do valor de 0,6. Dessa forma, pode-se inferir haver uma relação entre tais atributos, conforme foi elaborada essa análise de inferência estatística.

Para o caso do ajuste do modelo de Jacarta com R^2 com valor inferior a 0,5, foram gerados mapas de agrupamento LISA Moran's I para o atributo de energia e bivariado para os dois atributos de forma a observar do fenômeno.

Figura 6.36 Mapa de agrupamentos LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Jacarta

Fonte: o próprio autor do texto.

Figura 6.37 Mapa de índice de significância dos agrupamentos LISA Moran's I para atributo energia média consumida em viagens com destino na célula, Jacarta

Fonte: o próprio autor do texto.

Figura 6.38 Gráfico do indicador LISA Moran's I para os agrupamentos do atributo energia média consumida em viagens com destino na célula, Jacarta

Fonte: o próprio autor do texto.

Figura 6.39 Mapa de agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Jacarta

Fonte: o próprio autor do texto.

Figura 6.40 Mapa de índice de significância dos agrupamentos LISA Moran's I bivariado para os atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Jacarta

Fonte: o próprio autor do texto.

Figura 6.41 Gráfico do indicador LISA Moran's I bivariado para os agrupamentos dos atributos energia média consumida em viagens com destino na célula e quantidade de locais por célula, Jacarta

Fonte: o próprio autor do texto.

Da análise de todos os mapas de agrupamentos LISA Moran's I da cidade de Jacarta, incluindo o do atributo quantidade de locais, linhas acima, pode-se observar um fenômeno semelhante ao observado em Nova Iorque. Em Jacarta, onde aparecem alguns agrupamentos de locais, ainda que pequenos, na região central da cidade, há uma região de vazio ou não significativo agrupamento do atributo energia.

Tal fato suscita a ideia de que o uso naquela região de modo de transporte de baixo consumo unitário, motocicletas, linhas de ônibus e o transporte rápido de ônibus (BRT) lá existentes, impedem o surgimento de agrupamentos de energia do tipo alto-alto. Essa semelhança entre os mapas de agrupamentos das cidades de Nova Iorque e Jacarta, e os coeficientes de determinação dos modelos ajustados e apresentados linhas acima, sugerem a ideia de que em havendo a facilidade do modo de transporte de baixo consumo de energia nos locais de alta densidade de locais, e essa facilidade ser adotada por parcela significativa da mobilidade urbana daquela cidade, a relação entre a energia média consumida para se chegar às regiões de alta densidade de locais não se dê de forma consistente. Ou seja, a relação entre o consumo de energia para se locomover às regiões de alta densidade de locais de interesse em uma cidade pode depender da eficiência do sistema de transporte de baixo consumo unitário, os transportes rápidos de massa (MRT), metrô e ônibus comum ou sua versão aprimorada, o BRT.

A Figura 6.42, a seguir, apresenta o metrô de Nova Iorque na região enfocada, Manhattan, Bronx, Brooklyn e Queens, sobre o mapa de agrupamento LISA Moran's I do atributo energia. Nela pode-se observar que a área de não significância para agrupamentos do atributo energia (neste mapa e no da Figura 6.12, linhas acima), a qual não possui nenhum agrupamento alto-alto, é justamente a área de maior concentração de linhas de metrô naquela cidade, Manhattan e a ponta do Brooklyn junto de Manhattan.

Figura 6.42 Mapa de agrupamento LISA Moran's I do atributo energia média consumida em viagens com destino nas células com sistema de metrô dos bairros enfocados de Nova Iorque

Fonte: o próprio autor do texto.

7 CONCLUSÃO

7.1 Resultados e discussão

Este estudo:

- empregou dados desagregados das redes sociais a respeito de deslocamentos intraurbanos realizados de oito cidades mundialmente espalhadas e dados de duas pesquisas origem-destino, de Londres e São Paulo;
- utilizou algoritmos de Aprendizado de Máquina para atribuição do modo de viagem aos deslocamentos coletados das redes sociais para o cálculo do consumo de energia de cada deslocamento/viagem realizada, de forma a se obter qual cidade expressa o menor consumo de energia no funcionamento de sua mobilidade urbana;
- e, empregou técnicas de análise espacial com o intuito de se obter características ou diretrizes sustentáveis observadas na melhor prática, ou seja, na cidade de menor consumo de energia devido a sua mobilidade urbana.

Assim, buscou-se trazer contribuição à área de pesquisa que se ocupa da relação entre a energia e a mobilidade urbana. A novidade trazida advém de duas maneiras: a própria abordagem em si do tema, com uso de dados desagregados de deslocamentos intraurbanos, coletados a partir das redes sociais, com distâncias de viagens coletadas remotamente através de plataforma digital da *Internet*, a *Google Maps API*; e os resultados encontrados durante o desenvolver deste estudo.

A dificuldade para obtenção em escala mundial das pesquisas origem-destino e a demora na atualização de seus dados - pode chegar até uma década entre atualizações - fazem com que uma alternativa de coleta de deslocamentos interurbanos ágil e com abrangência mundial seja desejável. Embora haja viés no conjunto de dados das redes sociais, especialmente à época que foram coletados, 2011, época mais próxima do início do serviço da plataforma de *Internet*, donde originalmente os dados são provenientes, apenas uma determinada parcela da

população publicou seus deslocamentos. E, também, o uso do serviço foi percebido como variado no conjunto de dados entre as cidades, com alguns milhões de viagens em algumas cidades, e apenas algumas dezenas de milhares em outras. De qualquer forma, os resultados do método desenvolvido neste estudo demonstram a possibilidade de uso para fins de observação da mobilidade urbana de forma mundial.

A respeito do fato da demora para atualização das pesquisas origem-destino que impede uma observação ágil de fenômenos relativos a interferência no espaço urbano, os dados de deslocamentos provenientes das redes sociais podem vir a oferecer uma solução para essa lacuna temporal, uma vez que estão permanentemente disponíveis à coleta e com abrangência mundial, permitindo assim análises espaço-temporais impensáveis antes através de dados de seção cruzada das pesquisas origem-destino.

Os resultados obtidos na fase de modelagem de escolha do modo de viagem - ainda que o conjunto de dados das redes sociais somente contenha atributos de viagem -, com uso de algoritmos de aprendizado de máquina, obtiveram performance satisfatória na tarefa de classificação para quatro classes, a qual foi aproximadamente 70% para as duas cidades, Londres e São Paulo, onde se obteve as pesquisas origem destino para aferição. Essa boa performance é provavelmente devida à qualidade dos valores dos atributos coletados através da plataforma *Google Maps API*, com informações precisas sobre as distâncias de viagens, todos os transbordos e modos de viagens empregados no deslocamento total, incluindo uma precisa estimativa da duração da viagem, que leva em conta os horários de atendimento nos pontos das linhas de transporte público.

A translação espacial do modelo de escolha, ou seja, um modelo de escolha ajustado para somente as cidades de Londres e São Paulo, as quais forneceram os dados para este estudo, com uso do aprendizado de máquina propositalmente enviesado conforme a distribuição modal publicada da cidade de aplicação do modelo, mostrou-se útil para se superar a falta de informação sobre qual modo de viagem foi o empregado nos deslocamentos intraurbanos - falta essa característica de dados provenientes das redes sociais - em cidades mundialmente espalhadas, para as quais não se tenha a respectiva pesquisa origem-destino com a informação do modo de viagem escolhido para se ajustar um modelo de escolha para uma cidade. Tal constatação se apoia na similaridade observada entre as curvas de distâncias das pesquisas o-d de São Paulo e Londres, e as curvas obtidas com a atribuição do modo de viagem aos deslocamentos provenientes das redes sociais.

Portanto, uma contribuição deste estudo é que o ferramental desenvolvido propicia a obtenção de bases de dados a respeito da mobilidade urbana similares às pesquisas o-d em cidades distantes onde não exista a pesquisa o-d ou tal não seja fornecido. Outra vantagem desse ferramental desenvolvido neste estudo frente às pesquisas origem-destino convencionais é a rapidez de obtenção e o baixo custo envolvido, o que pode vir a viabilizar diversos estudos de planejamento antes impensados.

O ranque de consumo energético de cidades devido à mobilidade urbana - um dos objetivos deste estudo - é em parte consistente com estudos anteriores da literatura da área ao, por exemplo, apontar a cidade de Los Angeles como a de maior consumo. Em geral, a partir de indicadores de consumo médio obtidos se pode notar que o consumo de energia devido à mobilidade urbana em muito se deve ao modo e viagem utilizado, uma vez que as diferenças de consumo unitário entre as alternativas de modos, metrô, carro, ônibus, chegam a variar em até dez vezes, por exemplo, entre metrô e carro – por exemplo vide Tabela 5.2 para os consumos energéticos unitários dos modos da cidade de Paris, conforme UITP (2001).

A investigação conduzida sobre a relação entre o consumo médio de energia em viagens com destino inserido em células da malha, e a quantidade de locais inseridos na mesma célula da malha, demonstrou haver tal relação, apesar da falta de uma amostragem planejada sobre o universo do atributo densidade de locais de uma cidade. Destaca-se que somente foi utilizado o conjunto de locais fornecidos conjuntamente com o conjunto dos *check-ins* para esse estudo, sendo esses subconjuntos de 4 meses de coleta, daquele coletado por NOULAS (2013). Ou seja, o atributo quantidade de locais de interesse por células utilizado neste estudo é esperado como uma aproximação da situação real de mesma natureza.

Nas cidades onde não se observa a ocorrência de agrupamentos do atributo energia média consumida, nem alto-alto tampouco baixo-baixo, mas sim região insignificante para agrupamentos desse atributo, como foi o caso de Nova Iorque e Jacarta, e possuem agrupamentos tipo alto-alto para o atributo quantidade de locais na mesma região, ou seja região de alta densidade de locais, os modelos ajustados para esses dois atributos obtiveram menor índice de determinação, aproximadamente 0.45, em relação às demais cidades com índices de determinação variando de 0.63 até 0,75. Os dois fatos possivelmente indicam a eficiência do sistema de transporte com baixo consumo energético unitário implantado nessas regiões de alta densidade de locais de interesse.

Como possíveis diretrizes sustentáveis observadas na cidade de menor consumo energético, Nova Iorque, a partir dos dados utilizados neste estudo e as análises conduzidas, aponta-se o monocentrismo de distribuição de locais de interesse, ou seja, cidade tipo monocêntrica, e a oferta da facilidade de transporte de baixo consumo energético unitário, tipo MRT, com distribuição ampla na área urbana, de forma a conduzir o viajante até a área de alta densidade de locais onde a rede de tal sistema de transporte tenha diversidade de linhas sobre a área de alta densidade de locais, como se observa na rede de metrô de Nova Iorque. Porém, apenas a questão policentrismo ou monocentrismo tomada isoladamente pode não guardar correspondência direta com o consumo energético devido à mobilidade urbana. Nesse sentido, a cidade do Rio de Janeiro, que apresentou situação de distribuição espacial de locais policêntrica na análise de agrupamentos LISA Moran's I conduzida, situação essa que na literatura é considerada como tendente ao alto consumo energético devido à mobilidade urbana, figura entre as três primeiras no ranque de consumo obtido neste estudo, em grande medida devido à grande parcela de sua mobilidade prevista - pelo modelo de escolha ajustado - para o modo ônibus, um modo de transporte de baixo consumo energético unitário.

Destaca-se que as duas primeiras cidades do ranque de consumo energético obtido, além de apresentarem grande parcela de sua mobilidade atribuída pelo modelo de escolha ajustado ao modo de transporte de baixo consumo unitário, metrô, tiveram outra grande parcela de sua mobilidade atribuída ao modo caminhada, ou seja, sem consumo energético motorizado algum. Considerando a velocidade e o alcance do modo caminhada, pode-se inferir que tal grande parcela da distribuição modal ao modo caminhada se deve à característica monocêntrica daquelas duas cidades Nova Iorque e Paris.

A observação sobre como se dá a distribuição do consumo de energia devido à mobilidade urbana sobre o mapa da cidade, distribuição essa obtida de forma desagregada e vinculada a reais deslocamentos observados no espaço urbano como o ferramental desenvolvido neste estudo, abre possibilidade a diversas análises de impactos e melhorias de implantação e/ou alteração de facilidades de transportes no ambiente construído. Pois, tanto é possível se observar onde os picos ou agrupamentos de alto consumo de energia ocorrem, com possibilidade de conhecimento sobre a origem de tais viagens, ou seja, onde implantar uma facilidade de transporte de baixo consumo energético e para atender quem, como também é possível se obter de forma ágil uma análise do mesmo local após a implantação de uma facilidade de transporte, devido às características inerentes do ferramental desenvolvido neste estudo – permanente disponibilidade de coleta, agilidade e baixo custo.

7.2 Recomendações

Recomenda-se a aplicação do método proposto e utilizado neste estudo a uma maior quantidade de cidades, ou seja, o aumento da abrangência de escopo à procura de outras diretrizes sustentáveis que possam emergir de outras cidades com baixo indicador de consumo energético devido à mobilidade urbana. Inicialmente este estudo pretendia abranger uma quantidade maior de cidades que foi restringida devido ao custo de coleta de distâncias de viagens. Embora o indicador de consumo energético de Nova Iorque tenha sido bem abaixo do observado nas demais cidades, este pode ainda ser um mínimo local, não global, enquanto uma busca mais abrangente conduzida em mesmos moldes de dados desagregados não seja realizada para a apresentação de contra resultados.

Também é recomendável para trabalhos futuros uma coleta planejada dos locais de interesse da cidade de forma a se obter uma amostra com controle estatístico do universo de locais existentes na cidade sob análise.

REFERÊNCIAS

ABDELWAHAB, W. M. Transferability of intercity disaggregate mode choice models in Canada. **Canadian Journal of Civil Engineering**, 1991, v.18, n.1, p.20-26.

ANSELIN, L. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In *Spatial Analytical Perspectives on GIS in Environmental and Socio-Economic Sciences*. London: Taylor and Francis, 1996, p. 111-125.

ANSELIN, L. **Exploring spatial data with GeoDaTM** : a workbook. Center for Spatially Integrated Social Science. Revised version, 2005, 226 p. Disponível em:<<http://www.csiss.org/>>. Acesso em:

ANSELIN, L.; REY, S. **Perspectives on spatial data analysis**. Berlin: Springer, 2009.

ATHERTON, T. J.; BEN-AKIVA, M. E. Transferability and updating of disaggregate travel demand models. **Transportation Research Record**, 1976, n.610, p.12-18.

AXHAUSEN, K. **Can we ever obtain the data we would like to have?** In: **THEORETICAL Foundations of Travel Choice Modeling**. Oxford: Pergamon, 1998. p.305–334.

BANISTER, D.; WATSON, S.; WOOD, C. Sustainable cities: transport, energy, and urban form. **Environment and Planning B: Planning and Design**, 1997, v.24, p.125-143.

BARTER, P. A. An international comparative perspective on urban transport and urban form in pacific asia: responses to the challenge of motorisation in dense cities. Perth, Murdoch University, PhD thesis, 1999.

BARTHELEMY, M. Spatial networks. **Physics Reports**, 2011, v.499, p.101.

BATTY, M. **Cities as complex systems**: scaling, interactions, networks, dynamics and urban morphologies. London: Centre for Advanced Spatial Analysis, 2008. Working paper. CASA Working Papers (131).

BEN-AKIVA, M. E. **Issues in transferring and updating travel-behavior models**. In: **NEW horizons in travel behavior research**. Lexington: Lexington Books, 1981. p. 665-686.

BERTAUD A. & S. MALPEZZI. The spatial distribution of population in 48 world cities: implications for economies in transition. University of Wisconsin, 2003.

BRAND, D.; CHESLOW, M. D. Spatial, temporal, and cultural transferability of travel-choice models. In: **NEW horizons in travel-behavior research**. Lexington: Lexington Books, 1981.

BREHENY, M. The Contradictions of the compact city: a review. In: BREHENY, M. (Ed.). **Sustainable development and urban form Pion**. London: [s.n.], 1992, p.138-159.

BREIMAN, L. Random forests. **Machine Learning**, 2001, v.45, n.1, p.5-32.

BRONSTONE, D.; GOLOB, T. F. The Impact of residential density on vehicle usage and energy consumption. **Journal of Urban Economics**, 2009, v.65, p.91-98.

CANTARELLA, G. E.; DE LUCA, S. Multilayer feedforward networks for transportation mode choice analysis: an analysis and a comparison with random utility models. **Transportation Research C: emergency Technology**, 2005, v.13, n.2, p.121–155.

CARUANA, R.; NICULESCU-MIZIL, A. An Empirical comparison of supervised learning algorithms using different performance metrics. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 23., 2006, Pittsburgh. **Proceedings...** [S.l.:s.n.], 2006. p.161-168.

CHEN, M. et al. The Global pattern of urbanization an economic growth: evidence from the last three decades. **Plos One**, 2014, v.9, n.8, e103799.

CLARK, C. Transport – maker and breaker of cities. **Town Planning Review**, 1958, v.28, n.4, p.237-250.

COTTRILL, C. D. et al. Future mobility survey: experience in developing a smartphone-based travel survey in Singapore. Transportation Research Board 92nd Annual Meeting, 2013.

DAPENA, A. R.; MEDINA, J. R. Empleo de redes neuronales para modelos de eleccion de opciones de transporte. In: CONGRESO DE INGENIERIA DEL TRANSPORTE, 4., 2000, Valencia. **Anais...** [S.l.:s.n.], 2000, p.1-10.

ECHENIQUE, M. H. et al. Growing cities sustainably. **Journal of the American Planning Association**, 2012, v.78, n.2, p.121-137.

ECOTEC. Reducing transport emissions through planning HMSO. London, 1993.

ETTEMA, D.; TIMMERMANS, H.; VEGHEL, L. V. Effects of data collection methods in travel and activity research. Prepared for European Institute of Retailing and Services Studies, Tech. Rep., 1996.

EVI . Global EV outlook 2016: beyond one million electric cars, 2016. Disponible en:<www.iea.org/publications/freepublications/publication/Global_EV_Outlook_2016.pdf>.

FENG, T.; TIMMERMANS, H. J. P. Transportation mode recognition using GPS and accelerometer data. **Transportation Research C**, 2013, v.37, p.118-130.

FISCHER, M. M., GETIS, A. Handbook of applied spatial analysis: software tools, methods and applications. In: FISCHER, M. G. (ed.). Berlin Heidelberg: Springer-Verlag, 2010. 811 p.

FLACH, P. **Machine learning** – the art and science of algorithms that make sense of data. New York: Cambridge University Press, 2012.

FORTIN, M. J.; DALE, M. Spatial analysis – a guide for ecology. New York: Cambridge University Press, 2005. 365 p.

FOUQUET, R. Trends in income and price elasticities of transport demand (1850-2010). **Energy Policy**, 2012, v.50, p.62-71.

FRANKS. KOPPELMAN & CHESTER G. WILMOT **Transferability Analysis of Disaggregate Choice Models**. Transportation Research Record 895, 1982

GLOBAL ENERGY ASSESSMENT. **Toward a sustainable future**. New York: Cambridge University Press; Luxemburg: International Institute for Applied Systems Analysis, 2012

GLOBESCAN & MRC MCLEAN HAZEL. **Megacity challenges: a stakeholder perspective**. Munich: Siemens AG, 2007.

GOLDEMBERG, J. **Energy: what everyone needs to Know**. Oxford: Oxford University Press, 2012.

GOMEZ-IBANEZ, J. A global view of automobile dependence (Book Review). **Journal of the American Planning Association**, 1991, v.57, p.342-346.

GONZALEZ-ESCRIVA, J. A.; SANCHEZ, T. R. Utilizacion de redes neuronales par ajustar modelos de eleccion modal en transporte interurbano de personas. In: CONGRESO DE INGENIERIA DEL TRANSPORTE, 4, 2000, Valencia. **Anais...** [S.l.:s.n.], p.273-280.

HALL et al. The {WEKA} data mining software: an update. **SIGKDD Explorations**, v.11, n.1, 2009, p.10-18.

HAYKIN, S. **Neural networks – a comprehensive foundation**. 2nded. Upper Saddle River: Prentice Hall, 1999.

HENGL, T. Finding the right pixel size. **Computers & Geosciences**, 2006, v.32, p. 1283–1298.

HOLDEN, E.; LINNERUD, K.; BANISTER D. Sustainable passenger transport: back to brundtland. **Transportation Research Part A**, 2013, v.54, p.67-77.

HUBERT, J. P.; TOINT, P. L. From average travel time budgets to daily travel time distributions: appraisal of two conjectures by Koelbl and Helbing and some consequences. **Journal Transportation Research Board**, 2007, v.1985, p.135–143.

INTERNATIONAL ENERGY AGENCY. **ENERGY BALANCES OF NON-OECD COUNTRIES** (2015 edition). Paris, 2015. www.iea.org/statistics/topics/energybalances/

IQBAL, M. S. et al. Development of origin-destination matrices using mobile phone call data. **Transportation Research Part C: emerging technologies**, Mar. 2014, v.40, p.63-74.

JARIYASUNANT, J. et al. **The Quantified traveler: using personal travel data to promote sustainable transport behavior**. UC Berkeley: University of California Transportation Center, 2011.

KARASMAA, N. Evaluation of transfer methods for spatial travel demand models. **Transportation Research Part A**, 2007, n.41, p. 411-427.

KENWORTHY, J. Total daily mobility patterns and their policy implications for forty-three global cities in 1995 and 2005. *World Transport Policy and Practice*, January 2014, volume 20.1, p.41-55.

KENWORTHY, J.; LAUBE, F. **The Millennium cities database for sustainable transport**. Brussels: International Association of Public Transport; Perth: Institute for Sustainability and Technology Policy, 2001. CD ROM database.

KENWORTHY, J.; LAUBE, F. Urban transport patterns in a global sample of cities & their linkages to transport infrastructure, land use, economics & environment. **World Transport Policy & Practice**, 2002, v.8, n.3, p.5-19.

KOELBL, R.; HELBING, D. Energy and scaling laws in human travel behavior. **New Journal Physics**, 2003, v.5, n.48, p.1-16.

KOPPELMAN, F. S.; WILMOT, C. G. The Effect of omission of variables on choice model transferability. **Transportation Research Part B**, 1985, v.208, n.3, p.205-213.

KOPPELMAN, F. S.; WILMOT, C. G. Transferability analysis of disaggregate choice models. **Transportation Research Record**, 1982, n.895, p.18-24.

KOPPELMAN, F. S.; PAS, E. I. Multidimensional choice model transferability. **Transportation Research Part B**, 1986, v.20B, n.4, p.321-330.

KOPPELMAN, F. S.; KUAH, G. K.; WILMOT, C. G. Transfer model updating with disaggregate data. **Transportation Research Record**, 1985, n.1037, p.102-107.

LE NÉCHET, F. Urban spatial structure, daily mobility and energy consumption: a study of 34 European cities. *Cybergeo*. 2012, 10.4000/cybergeo.24966.

LEFEVRE, B. (2009). Urban transport energy consumption: an analysis of the literature. **S.A.P.I.E.N.S Surveys and Perspectives Integrating Environment and Society 2.3**, 2009, v.2, n.3, p.1-17 - Cities and Climate Change.

LIU, C. **Exploring the influence of urban form on travel and energy consumption, using structural equation modeling**, 2012. Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College Park. Doctoral thesis. Disponível em:<<http://hdl.handle.net/1903/12714>>.

LIU, C.; SHEN, Q. An Empirical analysis of the influence of urban form on household travel and energy consumption. **Computers, Environment and Urban Systems**, 2011, v.35, p.347-357.

MAMUN, S.; SABBIR, A. Spatial transferability of mode choice model. **International Journal of Engineering Research and Development**, July 2012, v.2, n.2, p.65-71.

MARTINELLI, E. Extração de conhecimento de redes neurais artificiais. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Dissertação de Mestrado, 1999, p.104.

METRÔ. **Pesquisa de mobilidade da região metropolitana de São Paulo**: síntese das informações – pesquisa domiciliar – dezembro de 2013. São Paulo: Secretaria dos Transportes Metropolitanos, 2013.

MINDALI, O.; RAVEH, A.; SALOMON, I. Urban density and energy consumption: a new look at old statistics. **Transportation Research Part A: policy and practice**, 2004, v.38, p.143-162.

MITCHELL, T. **Machine learning**. New York: McGraw-Hill, 1997.

MOONS, E., WETS, G., AERTS, M. Nonlinear models for determining mode choice. In: **PROGRESS in artificial intelligence**. Berlin Heidelberg: Springer, 2007, p. 183–194.

MURPHY, K. P. **Machine learning**: a probabilistic perspective. New York: MIT Press, 2012.

NEUMAN, M. The compact city fallacy. **Journal of Planning Education and Research**, 2005, v.25, p.11-26.

NEWMAN, P.; KENWORTHY, J. Is there a role for physical planners?. **Journal of the American Planning Association**, 1992, v.58, p.353-621.

NEWMAN, P.; KENWORTHY, J. Evaluating the transport sector's contribution to greenhouse gas emissions and energy consumption. In: SALTER, R.; DHAR, S.; NEWMAN, P. (Ed.). **Technologies for climate change mitigation** - transport sector. Denmark: UNEP Riso Centre on Energy, Climate and Sustainable, 2011, p. 7-23.

NEWMAN, W. G.; KENTHWORHTY, J. R. Gasoline consumption and cities. **Journal of the American Planning Association**, 1989a, v.55, 2, p.4-55.

NEWMAN, W. G.; KENTHWORHTY, J. R. **Cities and Automobile Dependence: an international sourcebook**. Brookfield: Gower, 1989b.

NIJKAMP, P. et al. Modelling inter-urban transport flows in Italy: a comparison between neural network analysis and logit analysis. **Transportation Research C**, 1996, v.4, n.6, p.323-338.

NOULAS, A. Human urban mobility in location-based social networks: analysis, models and applications. University of Cambridge, PhD thesis, 2013, 116 p.

NOULAS, A.; MASCOLO, C.; FRIAS-MARTINEZ, E. Exploiting foursquare and cellular data to infer user activity in urban environments. Proceeding MDM '13 Proceedings of the 2013 IEEE 14th International Conference on Mobile Data Management - Volume 01 Pages 167-176.

NOULAS, A. et al. A Tale of many cities: universal patterns in human urban mobility. **PLoS One**, 2012, v.7, n.5, e37027.

OMRANI, H. et al. Prediction of individual travel mode with evidential neural network model. **Transport. Res. Record: J. Transport. Research Board**, 2013, n.2399, p.1–8.

OPENSHAW. The Modifiable areal unit problem. Norwich: Geo Books, 1984.

PARSHALL, L. et al. Modeling energy consumption and CO2 emissions at the urban scale: methodological challenges and insights from the United States. **Energy Policy**, 2010, v.38, n.9, p.4765-4782.

QIAN, Y. et al. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. **Remote Sens.**, 2015, v.7, p.153-168. Doi:10.3390/rs70100153.

QUINLAN, J. R. **C4.5: programs for machine learning**. San Mateo: Morgan Kaufman, 1993.

REDDY, S. et al. Using mobile phones to determine transportation modes. **ACM Trans. Sensor Netw**, 2010, v.6, n.2, Article 13. Doi:<http://doi.acm.org/10.1145/1689239.1689243>.

RUTTER, P.; KEIRSTEAD, J. A Brief history and possible future of urban energy systems. **Energy Policy**, 2012, v.50, p.72-80.

SAFI, H.; MESBAH, M.; FERREIRA, L. ATLAS Project – developing a mobile-based travel survey. 36th Australasian Transport Research Forum 2013 Proceedings 2 – 4, October 2013, Brisbane, Australia Publication website: <http://www.patrec.org/atrf.aspx>

SALTER, R.; DHAR, S.; NEWMAN, P. **Technologies for climate change mitigation: Transportation sector**. United Nations Environment Programme (UNEP) Risø Centre on Energy, Climate and Sustainable Development, 2013.

SANTOSO, D. S.; TSUNOKAWA, K. Comparison of updating techniques in transferability analysis of work trip mode choice models in developing countries. **Journal of Advanced Transportation**, 2010, n.44, p.89-102.

SANTOSO, D. S.; TSUNOKAWA, K. Transferability and updating analysis of mode choice model between two developing countries – predictive performance point of view. **Journal of the Eastern Asia Society for Transportation Studies**, 2005, v.6, p.173 - 185.

SETO, K. C.; GÜNERALP, B.; HUTYRA, L. R. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. **Proceedings of the National Academy of Sciences**, 2012, v.109, n.40, p.16083-16088.

SMITH, M. E. Design of small sample home interview travel surveys. **Transportation Research Record**, n.701, p.29-35.

STENNETH, L. et al. Transportation mode detection using mobile phones and GIS Information. Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2011, p.54-63.

STOPHER, P. R.; WILMOT, C. G. Work-Trip mode-choice models for South Africa. **Transportation Engineering Journal**, 1979, v.105, n.6, p.595-608.

STOUFFER, S. Intervening opportunities: a theory relating mobility and distance. *American Sociological Review*, 1940.

SUSILO, Y.O.; STEAD, D. Urban form and the trends of transportation emissions and energy consumption of commuters in the Netherlands. The 87th Annual Meeting of the Transportation Research Board, Washington, D. C., 2008.

TALVITIE, A.; KIRSHNER, D. Specification, transferability and the effect of data outliers in modeling the choice of mode in urban travel. **Transportation**, 1978, n.7, p.311-331.

TEEuw, W. B.; KOOLWAAIJ, J.; PEDDEMORS, A. User behaviour captured by mobile phones. In: **CONSTRUCTING Ambient Intelligence. Communications in Computer and Information Science**. 2011, v.277, p.81-90.

TOBLER W. A Computer movie simulating urban growth in the Detroit region. **Economic Geography**, 1970, v.46, p.234-240. Supplement.

TORTUM, A.; YAYLA, N.; GOKDAG, M. The Modeling of mode choices of intercity freight transportation with the artificial neural networks and adaptive neuro-fuzzy inference system. **Expert Systems with Applications**, 2009, v.36, p.6199–6217.

UITP. Mobility in cities database. **International Association of Public Transport**, Brussels, 2001.

UNITED NATIONS FRAMEWORK CONVENTION ON CLIMATE CHANGE. **Paris declaration on electro-mobility and climate change and call to action**. 2015. Disponível em:<<http://newsroom.unfccc.int/lpaa/transport/the-paris-declaration-on-electro-mobility-andclimate-change-and-call-to-action>>.

VAPNIK, V.; GOLOWICH, S.; SMOLA, A. Support vector method for function approximation, regression estimation, and signal processing. In: MOZER, M.; JORDAN, M.; PETSCHKE, T. (Ed.). **Advances in neural information processing systems 9**. Cambridge: MIT Press, 1997, p.281– 287.

VAN AUDENHOVE, F.-J. et al. The Future of urban mobility 2.0: imperatives to shape extended mobility ecosystems of tomorrow. Report prepared by Arthur D. Little for UITP; January 2014.

WANG, T.; CHEN, C.; MA, J. Mobile phone data as an alternative data source for travel behavior studies. TRB – Transportation Research Board of the National Academies. Annual Meeting Online, 2013.

WATSON, P. L.; WESTIN, R. B. Transferability of Disaggregate Mode Choice Models. *Regional Science and Urban Economics*, 1975, n. 5, p.227-249.

WORLD COMMISSION ON ENVIRONMENT AND DEVELOPMENT. Our Common future. Oxford: Oxford University Press, 1987.

WERMERSCH, F. **Uso de redes neurais artificiais para descoberta de conhecimento sobre a escolha do modo de viagem.** 2002. Dissertação (Mestrado) - Universidade de São Paulo, Escola de Engenharia de São Carlos, São Carlos, 2002.

WILKERSON, G.; KHALILI, R.; SCHMID, S. Urban mobility scaling: lessons from ‘little data’. International Workshop on Network Science for Communication Networks, IEEE NetSciCom 2014.

WILLIAMS, N.; ZANDER, S.; ARMITAGE, G. A Preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. Article in ACM SIGCOMM Computer Communication Review, October 2006, DOI: 10.1145/1163593.1163596 · Source: DBLP.

XIE, C.; LU, J.; PARKANY, E. Work travel mode choice modeling with data mining: decision trees and neural networks. **Transport. Research Record: J. Transport. Res. Board**, 2003, n.1854, p.50–61.

XU, D. et al. Transportation modes identification from mobile phone data using probabilistic models. In: TANG, J. et al. (Ed.). **Advanced data mining and applications.** Berlin Heidelberg: Springer, 2011. v.7121, p.359–371.

ANEXO A – Método ANSELIN(2005)

Fonte: ANSELIN (2005).

APÊNDICE A - Feições do sistema de transporte e os pontos origem-destino da amostra de viagens provenientes das redes sociais nas 8 cidades enfocadas

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

APÊNDICE B – Estatística descritiva dos atributos: energia média consumida nos destinos e quantidade de locais, ambos inseridos na malha de célula 200 metros

Nova Iorque, atributo energia media consumida com destinos inseridos na célula (Avgtripe)	Count : 15715 Skewness : 8.5165 Min : 0 Kurtosis : 148.26 Max : 194380 1-st Quartile : 0 Mean : 1994.3 Median : 0 Std. Dev. : 5756.9 3-rd Quartile : 1355.5
Nova Iorque, atributo quantidade de locais inseridos na célula (Q_LD)	Count : 15715 Skewness : 6.5487 Min : 0 Kurtosis : 58.808 Max : 268 1-st Quartile : 0 Mean : 5.4454 Median : 1 Std. Dev. : 16.119 3-rd Quartile : 4
Paris, atributo energia media consumida com destinos inseridos na célula (Avgtripe)	Count : 36226 Skewness : 9.1582 Min : 0 Kurtosis : 113.41 Max : 78 1-st Quartile : 0 Mean : 0.73814 Median : 0 Std. Dev. : 3.5006 3-rd Quartile : 0
Paris, atributo quantidade de locais inseridos na célula (Q_LD)	Count : 36226 Skewness : 9.1582 Min : 0 Kurtosis : 113.41 Max : 78 1-st Quartile : 0 Mean : 0.73814 Median : 0 Std. Dev. : 3.5006 3-rd Quartile : 0
Rio de Janeiro, atributo energia media consumida com destinos inseridos na célula (Avgtripe)	Count : 29515 Skewness : 8.3835 Min : 0 Kurtosis : 100.44 Max : 182050 1-st Quartile : 0 Mean : 2024.6 Median : 0 Std. Dev. : 9153.9 3-rd Quartile : 0
Rio de Janeiro, atributo quantidade de locais inseridos na célula (Q_LD)	Count : 29515 Skewness : 13.664 Min : 0 Kurtosis : 262.92 Max : 140 1-st Quartile : 0 Mean : 0.83286 Median : 0 Std. Dev. : 4.566 3-rd Quartile : 0
Toronto atributo energia media consumida com destinos inseridos na célula (Avgtripe)	Count : 15170 Skewness : 4.4068 Min : 0 Kurtosis : 26.612 Max : 167850 1-st Quartile : 0 Mean : 4631.1 Median : 0 Std. Dev. : 14359 3-rd Quartile : 0
Toronto, atributo quantidade de locais inseridos na célula (Q_LD)	Count : 15170 Skewness : 9.4298 Min : 0 Kurtosis : 127.2 Max : 120 1-st Quartile : 0 Mean : 1.2334 Median : 0 Std. Dev. : 5.1684 3-rd Quartile : 1
Jacarta, atributo energia media consumida com destinos inseridos na célula (Avgtripe)	Count : 21405 Skewness : 3.4097 Min : 0 Kurtosis : 18.38 Max : 178940 1-st Quartile : 0 Mean : 6882.9 Median : 0 Std. Dev. : 15331 3-rd Quartile : 8021.8
Jacarta, atributo quantidade de locais inseridos na célula (Q_LD)	Count : 21405 Skewness : 22.686 Min : 0 Kurtosis : 854.44 Max : 889 1-st Quartile : 0 Mean : 2.1906 Median : 0 Std. Dev. : 16.389 3-rd Quartile : 0

Londres, atributo energia media consumida com destinos inseridos na célula (Avgtripe)	<table><tr><td>Count</td><td>: 51568</td><td>Skewness</td><td>: 8.8174</td></tr><tr><td>Min</td><td>: 0</td><td>Kurtosis</td><td>: 142.6</td></tr><tr><td>Max</td><td>: 430030</td><td>1-st Quartile</td><td>: 0</td></tr><tr><td>Mean</td><td>: 3265.5</td><td>Median</td><td>: 0</td></tr><tr><td>Std. Dev.</td><td>: 14117</td><td>3-rd Quartile</td><td>: 0</td></tr></table>	Count	: 51568	Skewness	: 8.8174	Min	: 0	Kurtosis	: 142.6	Max	: 430030	1-st Quartile	: 0	Mean	: 3265.5	Median	: 0	Std. Dev.	: 14117	3-rd Quartile	: 0
Count	: 51568	Skewness	: 8.8174																		
Min	: 0	Kurtosis	: 142.6																		
Max	: 430030	1-st Quartile	: 0																		
Mean	: 3265.5	Median	: 0																		
Std. Dev.	: 14117	3-rd Quartile	: 0																		

Fonte: o próprio autor do texto.

APÊNDICE C – Histogramas de distâncias de viagens por modos motorizados de São Paulo e Londres: pesquisas origem-destino e viagens das redes sociais

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.

Fonte: o próprio autor do texto.