ISSN: 1984-8781 - Anais XVIII ENANPUR 2019. Código verificador: **dwyXz73wNinL** verificar autenticidade em: http://anpur.org.br/xviiienanpur/anais

PREVISÃO DE MORBIDADE EM CIDADES Um estudo da poluição atmosférica e doenças respiratórias na Região Metropolitana de Curitiba (RMC)

Autores:
Fabio Teodoro de Souza - Pontifícia Universidade Católica do PR - fabio.teodoro@pucpr.br

Resumo:

Nas últimas duas décadas, a urbanização se intensificou, e no Brasil, cerca de 90% da população agora vive em centros urbanos. Padrões atmosféricos têm mudou devido à alta taxa de crescimento das cidades, com consequências negativas para a saúde pública. Essa pesquisa visa elucidar os padrões espaciais da poluição atmosférica e doenças respiratórias. Um modelo baseado em dados para ajudar gestão urbana para melhorar as políticas de saúde pública sobre a poluição do ar é descrito. Um exemplo de preparação de dados e análise multivariada com inventários de diferentes cidades da Região Metropolitana de Curitiba foi estudado. Um modelo preditivo com excelente precisão na previsão de surtos foi desenvolvida. A previsão da morbidade é uma estratégia sustentável para minimizar os efeitos dos surtos dessas doenças, pois aumenta o período de preparação das equipes de saúde, permite a promoção de campanhas de prevenção, otimiza os custos médicos e hospitalares, e elucida os padrões de ocupação das cidades pertinentes à gestão urbana. Os resultados preliminares descrevem relações relevantes entre escalas de morbidade, níveis de poluição do ar e padrões sazonais. O conhecimento reunido aqui contribui ao debate sobre questões sociais e políticas públicas. Além disso, os resultados deste estudo de menor escala podem ser estendido às megacidades

PREVISÃO DE MORBIDADE EM CIDADES

Um estudo da poluição atmosférica e doenças respiratórias na Região Metropolitana de Curitiba (RMC)

INTRODUÇÃO

A sustentabilidade de grandes cidades é influenciada pelo balanço de consumo, descarte e capacidade ambiental (Davis & Bardet, 2011). Um desafio atual é o entendimento dos padrões complexos dos efeitos da poluição do ar no meio ambiente e saúde humana e; a proposta de soluções sustentáveis para a mitigação dos efeitos adversos (Gurjar et al., 2010). Observações de constituintes atmosféricos relacionados à combustão têm sido utilizadas para caracterizar as emissões provenientes de ações antrópicas (Silva et al., 2013). O ciclo de carbono e sua interação com o clima na escala urbana têm sido investigados em estudos de sustentabilidade urbana.

A integração das ciências naturais e sociais tem sido proposto como um meio de compreender as mudanças na ecologia urbana em múltiplas escalas (Grimm et al., 2008). Uma revisão de literatura demonstrou que a sustentabilidade das cidades é ameaçada por processos metabólicos, tais como o fenômeno da ilha de calor do verão e armazenamento de calor na camada de dossel urbano. Estratégias de política urbana que consideram tanto o esgotamento de recursos e a desaceleração da exploração foram discutidos (Kennedy at el., 2007). Para entender o metabolismo urbano, comparações entre os ecossistemas selvagens e urbanos foram propostas e identificada a importância da análise das tendências da urbanização global nas megacidades (Decker et al., 2000).

A integração entre gestão urbana e saúde pública é uma necessidade para se encontrar soluções aos diversos problemas da sociedade em densos ambientes urbanos. A poluição do ar resultante das emissões de gases de indústrias e especialmente dos veículos motorizados tem impacto negativo na saúde respiratória da população. Dentre os estudos de integração, uma abordagem para analisar a associação entre características do ambiente construído e morbidade relacionada a doenças crônicas não transmissíveis é discutida (Hino et al., 2013). O transporte público e os impactos negativos na saúde pública também são analisados (Mosquera et al., 2013). O compreendimento desses padrões de causa e efeito contribui para o processo de tomada de decisão dos gestores quanto às práticas de saúde e medidas de prevenção de doenças (Brownson et al., 2014).

Um estudo identificou associação entre os dados diários de poluição atmosférica e atendimentos de pneumonia e gripe para idosos entre 1996 e 1998 (Martins et al., 2002).

Bakonyi et al. (2004) investigaram relações qualitativas entre morbidade respiratória da população infantil e poluentes atmosféricos em Curitiba (cuja área metropolitana possui população aproximada de 3 milhões de pessoas).

A extensão de metodologias de cidades com tamanho médio para ambientes mais complexos e representativos das megacidades (áreas metropolitanas com população superior a 10 milhões) é uma necessidade enfatizada na literatura (Duren & Miller, 2012). Devido ao expressivo crescimento nas emissões de dióxido de carbono pela queima de combustíveis fósseis, os autores ressaltam a urgência da implantação de sistemas de monitoramento nas megacidades para dar suporte às decisões sobre as políticas de controle do carbono urbano. Também é enfatizada a importância de monitorar continuamente dióxido de carbono, metano e monóxido de carbono, juntamente com a análise cientificamente relevante e compartilhamento de dados entre as cidades de diferentes países. Tais medidas contribuiriam para uma melhor eficácia nas políticas de mitigação de emissões.

Cientistas discutiram a dificuldade de avaliar o efeito da poluição do ar nas megacidades dos países em desenvolvimento por causa dos pobres sistemas de monitoramento ambiental, da falta de políticas coesas de qualidade do ar e da escassez de dados de vigilância de doenças (Shah et al., 2013). Concluem que a poluição do ar tem uma direta associação com hospitalização e mortalidade por insuficiência cardíaca e destacam que a poluição do ar deve ser objetivo fundamental para as políticas de saúde global.

Um estudo descreve o uso de imagens de satélite das luzes noturnas para avaliar as emissões de dióxido de carbono provenientes do consumo de energia residencial nas cidades (Lu & Liu, 2014). Ressaltam os diversos impactos na saúde devido à poluição atmosférica associada ao tráfego e propõem o rastreamento por telefonia móvel (Liu et al., 2013).

Este artigo descreve uma pesquisa científica que se utiliza de uma metodologia de monitoramento de padrões adversos de qualidade do ar nos centros urbanos e das consequências desfavoráveis à saúde pública. A identificação dos padrões de relação entre causa e efeito pode revelar informações de gestão urbana características para as diferentes manchas urbanas, ou seja, serve como diretriz aos planos de ação regionais. A pesquisa propõe ainda a previsão de morbidade em sete cidades da Região Metropolitana de Curitiba. E como perspectiva futura, a pesquisa ainda pretende explorar as relações de fenômenos ambientais urbanos e o monitoramento de CO2 em megacidades.

FUNDAMENTAÇÃO TEÓRICA

Muitos inventários de emissões de CO2 são descritos na literatura. As discrepâncias entre os inventários nos diferentes níveis de governança com erro equivalente de 5% do total global de emissões são discutidas (Duren & Miller, 2012; Guan et al., 2012). Os métodos para a redução de escala de inventários nacionais para escalas espaciais e temporais menores têm sido desenvolvidos, mas tais métodos introduzem incertezas significativas e divergências numéricas.

O inventário denominado "Vulcan" contém dados de emissões de CO2 pela combustão de fósseis nos EUA com escala espacial aproximada de 100 km^2 e escala temporal horária. Este inventário tem expressiva relevância na elucidação do ciclo do carbono e mudanças climáticas em megacidades. Por exemplo, o CO2 de combustíveis fósseis desempenha um papel fundamental nas inversões de CO2 na atmosfera, padrões amplamente utilizados para a compreensão do fluxo líquido terra-atmosfera.

Os resultados iniciais do monitoramento de CO2 na megacidade de Los Angeles e comparações com níveis de gás de efeito estufa em regiões urbanas são debatidos (Newman et al., 2013). As emissões foram rastreadas por meio de instrumentos de sensoriamento remoto espaciais de alta resolução, como o Observatório de Carbono Orbital (OCO-2, resolução espacial de aproximadamente 3 km^2). Com o objetivo de caracterizar o sinal antropogênico e elucidar os padrões de emissões urbanas foram também realizadas medições *in situ* em vários locais.

Uma metodologia para quantificar as emissões de CO2 de combustíveis fósseis em toda a cidade de Indianápolis (EUA) é apresentada (Gurney et al., 2012). As emissões são quantificadas com resolução espacial de edifícios individuais, segmentos de estrada, e instalações de produção industrial / eletricidade; e com resolução temporal horária. Chamado de projeto "Hestia", a pesquisa combina uma série de conjuntos de dados em energia de construção, tráfego, produção de energia, e poluição do ar local.

A importância de acompanhar emissões de gases de efeito estufa provenientes da queima antrópica e o impacto na mudança climática de longo prazo é ressaltada (Silva et al., 2013). O estudo usou dados de sensoriamento remoto (satélites GOSAT e MOPITT) e comparou a escalabilidade espacial entre megacidades e países. Kort et al. (2012) discutiram o potencial da utilização do GOSAT para avaliar as emissões de CO2 em Los Angeles e Mumbai.

O projeto MEGAPOLI enfoca os impactos da poluição extrema do ar em megacidades (Von Der Weiden-Reinmüller et al., 2014). Os parâmetros medidos na área metropolitana de Paris (França) incluem concentrações da fase gasosa de ozônio, NOx and CO2, concentração de partículas de aerossol, distribuição de tamanho de partículas e composição química, bem como variáveis meteorológicas (direção do vento, umidade relativa, pressão e temperatura).

A maior parte da poluição de Deli (Índia) é devido às atividades de construção, motorização, e geração de energia (Guttikunda & Calori, 2013). Os autores descrevem um inventário de emissões de CO, NOx, PM, SO2, e compostos orgânicos voláteis. Ainda são propostos modelos de dispersão de partículas para comparar e validar o inventário de emissões (com resolução espacial de pixels de 1×1 km e resolução temporal de 1h).

O aumento expressivo das emissões de gases de efeito estufa e de poluentes no ar na China é relacionada com a expansão econômica do país. Grande parte da expansão é impulsionada pelo uso de combustíveis fósseis (Kan et al., 2012). São discutidas as associações significativas entre os níveis de poluição do ar e mortalidade por doenças cardiopulmonares e câncer no pulmão.

A presente pesquisa vislumbra extrair conhecimento sobre a gestão urbana no contexto de poluição atmosférica e doenças respiratórias de diferentes fontes de informações, sejam de material bibliográfico, dados reais e históricos sobre a temática de doenças respiratórias em grandes centros urbanos, e de especialistas nas áreas de planejamento urbano, medicina, mudanças climáticas, ciências ambientais, física e química atmosférica, geofísica, entre outras que forem pertinentes. A extração de conhecimento proposta nesse projeto considera a aplicação de técnicas em mineração de dados.

MÉTODO DE PESQUISA

Um projeto que envolve mineração de dados é usualmente desenvolvido em três etapas: coleta de dados, preparação dos dados e modelagem.

Coleta de Dados

A metodologia utilizada centra-se na análise de dados reais através da aplicação de técnicas de mineração de dados e sistemas de informação geográfica, bem como a utilização de conhecimento especialista em modelos computacionais. Um estudo preliminar incluiu uma análise da qualidade do ar e dados de morbidade entre 2008 e 2016 nas cidades da Região Metropolitana de Curitiba (RMC): Almirante Tamandaré, Araucária, Campo Largo, Colombo, Curitiba, Pinhais e Piraquara. A seleção de tais cidades resulta primeiro pelas adiantadas relações interinstitucionais das agências governamentais atuantes nestas cidades. A escolha das seis cidades circunvizinhas ao município principal de Curitiba também considera a abrangência da amplitude de direção do vento, pois as cidades vizinhas contemplam grande parte dos 360 graus de azimute. A seleção de Araucária diz respeito, sobretudo à sua concentração de plantas industriais e já conhecidos problemas respiratórios devido à emissão de poluentes da importante cidade industrial. A seleção de Colombo diz respeito ao processo de industrialização que o município tem se submetido nas duas últimas décadas, onde se constitui um novo polo industrial na RMC, e também de significativa concentração de população de baixa renda territorial.

O IAP (2014) monitora sete parâmetros de qualidade do ar e disponibiliza no portal eletrônico medições diárias de partículas totais em suspensão (PTS); fumaça; partículas inaláveis (PI ou PM10); dióxido de enxofre (SO2); monóxido de carbono (CO); ozônio (O3); dióxido de nitrogênio (NO2). Além dos parâmetros medidos, o IAP também compõe um índice de qualidade do ar (IQA) e categoriza o ar e o impacto na saúde humana (Tabela 1). Infelizmente ainda não existe o monitoramento de CO2 em Curitiba.

O presente estudo também relaciona a poluição do ar à morbidade e variáveis climáticas. Uma base de dados foi obtida de uma estação meteorológica localizada em Curitiba e monitorada pelo Instituto Nacional de Meteorologia (INMET). Os dados contemplam registros diários de temperatura, pressão atmosférica, velocidade e direção do vento e umidade relativa. Esses parâmetros foram relacionados aos índices sazonais de poluição do ar e doenças respiratórias (Bakonyi et al., 2004). Duchiade (1992) investigou a mortalidade

devido a poluentes atmosféricos e a relação de vários parâmetros atmosféricos, tais como precipitação, pressão atmosférica, temperatura, umidade, vento e radiação. A incidência de doenças respiratórias em crianças e idosos em Campina Grande foi relacionada com variáveis climatológicas (temperatura, umidade relativa do ar e pluviosidade) e com variáveis do processo de urbanização no município (De Moura, 2009).

Tabela 1. Índices de qualidade do ar e impactos na saúde humana

Qualidade do AR (IQA)	**Categoria**	**Impacto**
Até 50	BOA	Nenhum ou muito pequeno
51 e 100	REGULAR	Apenas em pessoas muito sensíveis
101 e 150	INADEQUADA	Em pessoas sensíveis
151 e 200	INADEQUADA	Em pessoas com sensibilidade média, de forma mais acentuada em pessoas sensíveis
201 e 300	MÁ	Em pessoas com sensibilidade média e com efeitos mais graves
> 300	PÉSSIMA ou CRÍTICA	Na população em geral

Fonte: Instituto Ambiental do Paraná.

Principalmente, estão sendo incluídos nesse estudo os dados de internações hospitalares por doenças respiratórias tanto em âmbito ambulatorial quanto clínico-hospitalar e em diferentes níveis de obtenção: Federal (DATASUS), Estadual (SESA) e das secretarias municipais de Curitiba, Araucária e Colombo.

Além disso, a caracterização das cidades com relação ao uso e ocupação do solo também será incluída no conjunto de dados, seja por meio do cálculo da proximidade de parques industriais e atividades de mineração, e pelas características espaciais de veículos a motor (zonas de intensa circulação).

A alta concentração de CO2 pode retardar o efeito da mudança climática, mas com reflexo ambiental local em curto prazo (D'Angiola et al., 2010). Uma metodologia propõe análise de um inventário anual de emissões como uma ferramenta de gestão ambiental para identificar (i) a relação entre um indicador socioeconômico e número de veículos, e (ii) defasagens temporais em tecnologias de veículos entre países em desenvolvimento e desenvolvidos. Estratégias em mineração de dados poderiam ser adaptadas para estimar o tempo de retardo do efeito na mudança climática ou de explicar o impacto das emissões de CO2 provenientes de combustível fóssil em mudança climática. O monitoramento de CO2 da RMC também é um dos objetivos desta pesquisa em passos futuros.

Preparação de Dados

Esta etapa é a mais importante de todo o processo e a que consome a maior parte do tempo (Pyle, 1999). Nessa etapa devem ser corrigidas as inconsistências, como por exemplo, falsos valores de zero, tratados os *outliers*, preenchidas as falhas (quando possível), entre outras medidas para preparar os dados para a modelagem. Uma vez que os dados são consistidos, então os mesmos devem ser inseridos em um sistema georreferenciado (SIG).

Autores descrevem uma metodologia para integrar dados de inventários múltiplos e heterogêneos e aperfeiçoar a inferência de resolução espaço-temporal (Braverman, 2012). A metodologia detalha a integração dos dados de sensoriamento remoto dos satélites AIRS e GOSAT.

A manipulação preliminar dos dados consistidos em uma plataforma georreferenciada permite a visualização gráfica com clareza da relação entre as variáveis envolvidas e a elucidação de alguns padrões que auxiliam na estratégia de modelagem.

Modelos geo-estatísticos foram utilizados para a detecção de "*hot spots*" ou zonas de riscos elevados à saúde, devido à má qualidade do ar e as altas temperaturas em Aachen, Alemanha (Merbitz et al., 2012). Estudo semelhante estabeleceu relações entre exposição ao ar poluído e mortalidade de pessoas, e a identificação de áreas de alto e baixo nível de exposição em Auckland, Nova Zelândia (Scoggins et al., 2004). Um modelo georrefenciado denominado ARMOS foi aplicado para simular a concentração de poluentes na área urbana de Hagzhou, China (Zhang et al., 2008). Um modelo georreferenciado foi integrado para calcular e visualizar geograficamente a distribuição de Monóxido de Carbono (CO) ao longo de um trecho de rua com edificações (Wang, 2005).

O impacto ao meio ambiente causado pela construção de uma autoestrada nas imediações de Madri na Espanha foi avaliado razoável acurácia (Moragues & Alcaide, 1996). A exposição de crianças em idade escolar a poluentes durante o tempo de trajeto entre suas casas e as escolas foi simulada por modelos georreferenciados (Gulliver & Briggs, 2005). Considera-se, portanto, que a contribuição do uso de técnicas em geoprocessamento para preparar os dados é relevante para esta pesquisa.

Este artigo descreve resultados preliminares de uma análise sazonal dos padrões atmosféricos e morbidade por doenças respiratórias. Foram analisados os padrões sazonais de variáveis climáticas e partículas inaláveis medidas em Curitiba, e a morbidade em Curitiba e Campo Largo, as duas cidades com maior percentual de morbidade das sete escolhidas.

Modelagem

A etapa de modelagem aborda duas fases: análise multivariada e modelos quantitativos. A análise multivariada consiste principalmente em identificar padrões qualitativos entre as variáveis envolvidas (Duarte et al., 2016). A identificação clara das relações entre as variáveis envolvidas orienta na construção dos modelos quantitativos (Pinzón & Souza, 2016).

Uma metodologia propõe uma combinação de dados de sensoriamento remoto e outras fontes de dados para estudar mudanças climáticas (Wentz et al., 2014). São discutidas abordagens analíticas avançadas, e incluem técnicas de mineração de dados para os desafios emergentes e urgentes que influenciam a mudança ambiental global.

O uso de técnicas de mineração de dados também tem sido considerado para explicar as mudanças rápidas na urbanização e população global (Faghmous & Kumar, 2014). As mudanças climáticas têm sido explicadas por modelos espaço-temporal de mineração de dados com aumento de precisão, escalabilidade ou interpretação (em relação aos algoritmos não-espaço-temporal).

Uma abordagem que permite a redução dos riscos ambientais globais com a aplicação de mineração de dados em ambientes de grande variabilidade espaço-temporal tem sido discutida (Morik et al., 2012). Tal abordagem poderia ser adaptada para projetar e operar sistemas mais sustentáveis referentes à poluição do ar.

Nesta pesquisa também estão sendo desenvolvidos modelos matemáticos para a previsão de morbidade nas sete cidades da RMC, com até quatro meses de antecedência. A previsão da morbidade é uma estratégia sustentável para minimizar os efeitos dos surtos dessas doenças, pois aumenta o período de preparação das equipes de saúde, permite a promoção de campanhas de prevenção, otimiza os custos médicos e hospitalares, e elucida os padrões de ocupação das cidades pertinentes à gestão urbana (Souza, 2018).

A etapa de modelos quantitativos aborda a geração de regras de associação do tipo: SE (A), ENTÃO (B). Uma regra é um modelo simples e de fácil interpretação que relaciona causa e efeito pela relação SE/ENTÃO (Agrawal & Srikant, 1994; Agrawal et al., 1993). Essas regras podem ter caráter preditivo e são chamadas regras de classificação (Liu et al., 1998). Portanto, a previsão pretende identificar três níveis de morbidade (*baixo*, *médio* e *alto*) para alertar as autoridades sobre a possível ocorrência de surtos de doenças respiratórias.

Um exemplo de regra de classificação gerada para prever deslizamentos de encostas durante fortes chuvas na cidade do Rio de Janeiro é descrita (Souza & Ebecken, 2012):

SE Chuva Acumulada nas últimas 6 horas for maior que 43.7mm (A)

ENTÃO Ocorrência de Deslizamento (B)

(9,24% % 90,6% 117 106)

Os itens percentuais abaixo da regra são os valores de suporte e confiança. O suporte da regra é 9,24% e é calculado pela divisão de 117 vezes em que (A) ocorreu pelo total de valores da amostra, ou os 1266 registros que incluem *deslizamentos*, *pânico*, ou *não ocorrência* (embora chova bastante). Dessas 177 vezes que (A) ocorreu no banco de dados, 106 vezes também ocorreram (B), e, portanto, a confiança da regra é de 90,6% (106 ⁄117*100). Observa-se que a regra é de fácil interpretação e acionável, ou seja, poderia ser

utilizada pelo poder público na emissão de alertas de deslizamentos pela simples leitura dos índices de chuva monitorados pelos pluviômetros automáticos distribuídos na cidade.

Outra grande vantagem da geração de regras, é que o conhecimento adquirido nos dados pode ser integrado com o conhecimento especialista (Liu et al., 2000). O estudo visa descobrir regras inesperadas aos especialistas e que possuam alto grau de "interesse". A combinação de informações considera entrevista aos especialistas nas diversas áreas envolvidas sobre poluição do ar e saúde pública.

RESULTADOS E DISCUSSÃO

Este artigo apresenta uma contribuição sobre a preparação de dados para a análise multivariada e modelagem, uma análise da prevalência sazonal da morbidade e a relação com a poluição do ar e climatologia, e de como gerar regras multivariadas para a previsão de morbidade urbana por doenças respiratórias.

Análise Multivariada

Este item descreve a metodologia usada para a cidade de Curitiba. Para encontrar correlações relevantes entre as variáveis, quatro técnicas de mineração dados foram aplicadas: duas estatísticas (análise fatorial e correlação linear) e duas de agrupamento de dados ou *clustering* (*k-means* e dendrograma).

O conjunto de dados é composto por totais mensais de pacientes internados por doenças respiratórias de janeiro de 2008 a agosto de 2016. A amostra contempla diversidade completa (de jovens a idosos, homens e mulheres etc.). A variável de morbidade de Curitiba foi assim rotulada MS_CWB_DR_T0; onde M é a morbidade, S é ao sul da área metropolitana, CWB denota Curitiba, DR é doença respiratória, e T0 é o mês atual. Desta variável, quatro novas variáveis de morbidade foram criadas para próximo (T1), segundo (T2), terceiro (T3) e quarto (T4) meses. Estas novas variáveis são a saída do modelo de previsão. Além disso, dados da qualidade do ar de duas estações em Curitiba (Santa Cândida e Ouvidor Pardinho) foram adicionados à matriz. Finalmente, algumas variáveis meteorológicas foram adicionadas à matriz. A matriz inteira consiste de 19 variáveis (15 variáveis explicativas e 4 variáveis de saída) e 104 instâncias dos registros mensais.

A correlação linear foi a primeira técnica aplicada. O valor *p* para realçar a significância é 0,05. A Tabela 2 mostra os coeficientes de correlação linear. Um coeficiente linear próximo de + 1 ou -1 representa uma correlação linear positiva ou negativa perfeita, respectivamente. Observa-se que a diagonal principal apresenta valores iguais a 1, pois a correlação linear de uma variável com ela mesma é perfeita. Valores destacado em cinza correspondem a valores maiores que + 0,57 ou menor que - 0,57. Este valor foi escolhido porque a melhor correlação linear de uma variável de saída foi de 0,58 (morbidade por doenças respiratórias nos próximos mês, M_S_CWB_DR_T1). As letras A até E na parte inferior da Tabela 2 são definidas da seguinte forma. Existe uma correlação linear inversa igual a - 0,64 entre "mês" e morbidade no terceiro e quarto meses (A) e pode haver um padrão sazonal de linearidade nesses meses.

Tabela 2. Coeficientes de correlação linear

<_-0,57_(Hachurado)_>_+0,57	Mês	Dir_V	Vel_Vento_méd_m/s	Pressão_Atm_hPa	Temp_Min_Med_ºC
Mês	1				
Dir_V	-0,09	1			
Vel_Vento_méd_m/s	0,20	-0,15	1		
Pressão_Atm_hPa	0,01	0,09	-0,38	1	
Temp_Min_Med_ºC	-0,29	-0,10	0,31	-0,80	1
Umidade_Méd_%	-0,24	0,09	-0,01	0,02	-0,02
SO2_STCND	-0,03	0,27	0,09	0,10	-0,04
O3_STCND	0,47	0,13	0,04	0,24	-0,38
NO2_STCND	-0,04	0,17	-0,54	0,63	-0,58
PTS_OUVPD	0,09	0,17	-0,53	0,71	-0,72
PI_OUVPD	0,17	0,15	-0,51	0,70	-0,73
SO2_OUVPD	0,17	-0,05	-0,23	0,33	-0,38
O3_OUVPD	0,50	0,02	0,18	-0,10	-0,01
NO2_OUVPD	0,04	0,12	-0,72	0,40	-0,41
M_S_CWB_DR_T0	0,34	0,15	-0,34	0,51	-0,69
M_S_CWB_DR_T1	-0,09	0,07	-0,48	0,58	-0,54
M_S_CWB_DR_T2	-0,56	0,04	-0,39	0,46	-0,28
M_S_CWB_DR_T3	-0,64	0,03	-0,38	0,13	0,12
M_S_CWB_DR_T4	-0,64	0,08	-0,12	-0,28	0,40
	A	B	C	D	E

Fonte: Autor, IAP, INMET, DATASUS.

Tabela 3. Coeficientes de carga fatorial

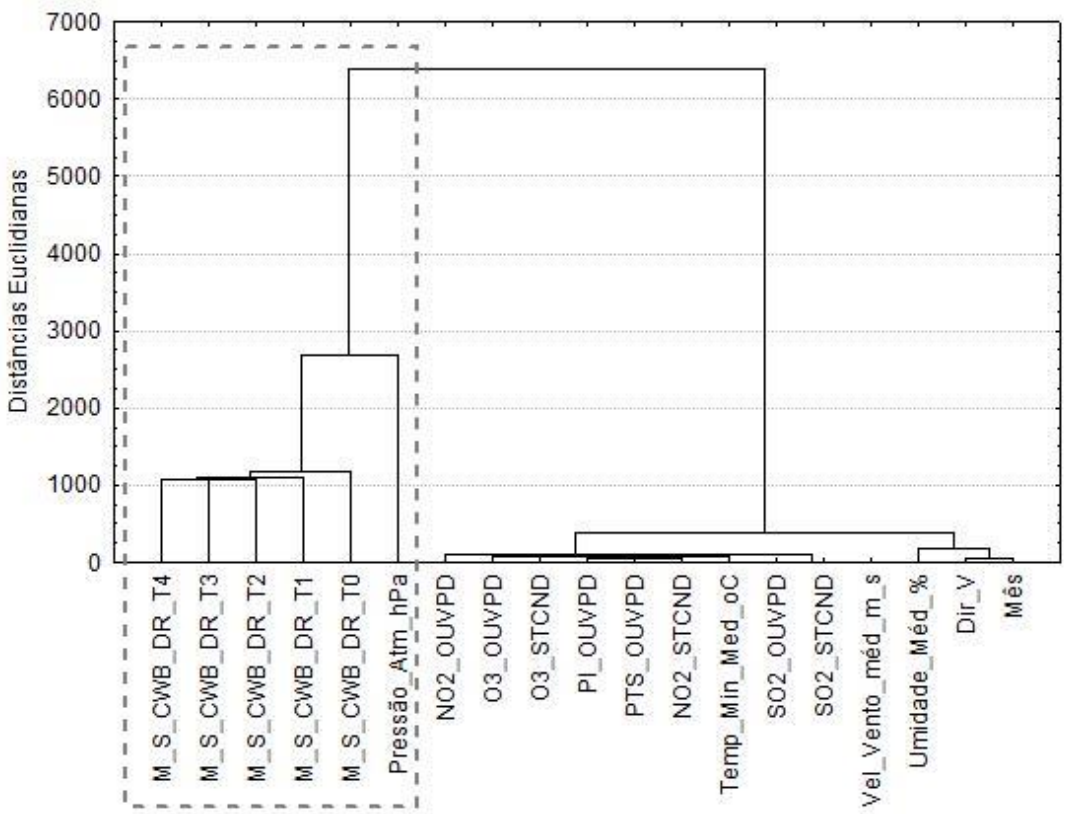
<_-0,57_(Hachurado)_>_+0,57 **Variável**	Fator 1	Fator 2	Fator 3	Fator 4
Mês	0,11	0,84	-0,12	0,15
Dir_V	0,17	-0,06	-0,16	-0,64
Vel_Vento_méd_m/s	-0,64	0,35	-0,27	0,06
Pressão_Atm_hPa	0,83	-0,04	-0,25	-0,05
Temp_Min_Med_ºC	-0,84	-0,21	0,32	-0,03
Umidade_Méd_%	-0,14	-0,41	-0,72	0,16
SO2_STCND	-0,07	0,05	-0,35	-0,78
O3_STCND	0,30	0,65	-0,07	0,01
NO2_STCND	0,77	-0,11	0,04	-0,18
PTS_OUVPD	0,89	0,05	0,21	-0,18
PI_OUVPD	0,90	0,13	0,22	-0,16
SO2_OUVPD	0,44	0,23	0,24	0,27
O3_OUVPD	0,01	0,71	0,29	-0,05
NO2_OUVPD	0,67	-0,06	0,38	-0,05
M_S_CWB_DR_T0	0,69	0,09	-0,38	0,23
M_S_CWB_DR_T1	0,72	-0,28	-0,20	0,29
M_S_CWB_DR_T2	0,47	-0,68	-0,13	0,15
M_S_CWB_DR_T3	0,15	-0,84	0,13	0,01
M_S_CWB_DR_T4	-0,25	-0,74	0,36	-0,07

Fonte: Autor, IAP, INMET, DATASUS.

No entanto, a maioria dos valores de coeficientes linear apresentam baixa correlação, como na coluna B (Direção do vento). Valores de correlação linear baixa não são discutidos. A coluna C apresenta uma correlação linear negativa entre velocidade do vento e concentração de dióxido de nitrogênio na Estação do Ouvidor Pardinho. A coluna D mostra que a pressão do ar apresenta correlação linear com concentração de poluentes atmosféricos no centro da cidade e com morbidade mensal atual (T0). A temperatura mínima média (E) também apresenta correlação linear relevante com concentração de poluentes atmosféricos no centro e com a morbidade do próximo mês (T1).

A análise fatorial processou 104 casos e uma matriz de correlação foi computada para 19 variáveis. O método de extração de componentes principais foi utilizado com número de fatores iguais a 4 e um autovalor mínimo igual a 1. A Tabela 3 ilustra as cargas fatoriais para as variáveis. Os valores destacados em cinza na Tabela 3 correspondem coeficientes superiores a + 0,57 ou inferiores a - 0,57 (o mesmo critério usado para correlação linear). Os resultados foram combinados a partir de cargas do fator 1 e da matriz de correlação linear para selecionar um conjunto de nove variáveis a serem inseridas para a construção do modelo de previsão (Modelo 1).

Figura 1. Dendrograma (Distância Euclidiana)



Fonte: Autor, IAP, INMET, DATASUS.

Duas técnicas de agrupamento foram aplicadas para encontrar relações entre entrada e saída. Um dendrograma é um resultado gráfico da análise hierárquica de "agrupamento de árvores". A Figura 1 mostra um dendrograma construído com a Distância Euclidiana. Observa-se dois "galhos" distintos. As variáveis de saída (morbidade para o primeiro, segundo, terceiro e quarto meses) são agrupadas com duas variáveis de entrada: morbidade no mês atual (T0) e pressão atmosférica (Pressão_Atm_hPa), conforme destaca o retângulo tracejado.

A técnica *k-means* foi aplicada considerando dois clusters, classificar a distância e fazer observações em intervalos constantes. A Tabela 4 mostra que o Cluster 1 agrupa as mesmas variáveis de saída de acordo com o dendrograma. Dessa forma, essas duas variáveis de entrada (morbidade no mês e pressão do ar) são usadas como entrada para o modelo preditivo (Modelo 2).

Tabela 4. Método k-means

Cluster_1		***Cluster_2***	
Variável	***Distância***	***Variável***	***Distância***
Pressão_Atm_hPa	269,1	Mês	71,2
M_S_CWB_DR_T0	155,0	Dir_V	67,1
M_S_CWB_DR_T1	118,8	Vel_Vento_méd_m/s	33,1
M_S_CWB_DR_T2	106,3	Temp_Min_Med_ºC	21,8
M_S_CWB_DR_T3	121,2	Umidade_Méd_%	45,9
M_S_CWB_DR_T4	152,0	SO2_STCND	33,9
		O3_STCND	12,2
		NO2_STCND	15,3
		PTS_OUVPD	13,4
		PI_OUVPD	18,4
		SO2_OUVPD	34,1
		O3_OUVPD	10,4
		NO2_OUVPD	11,1

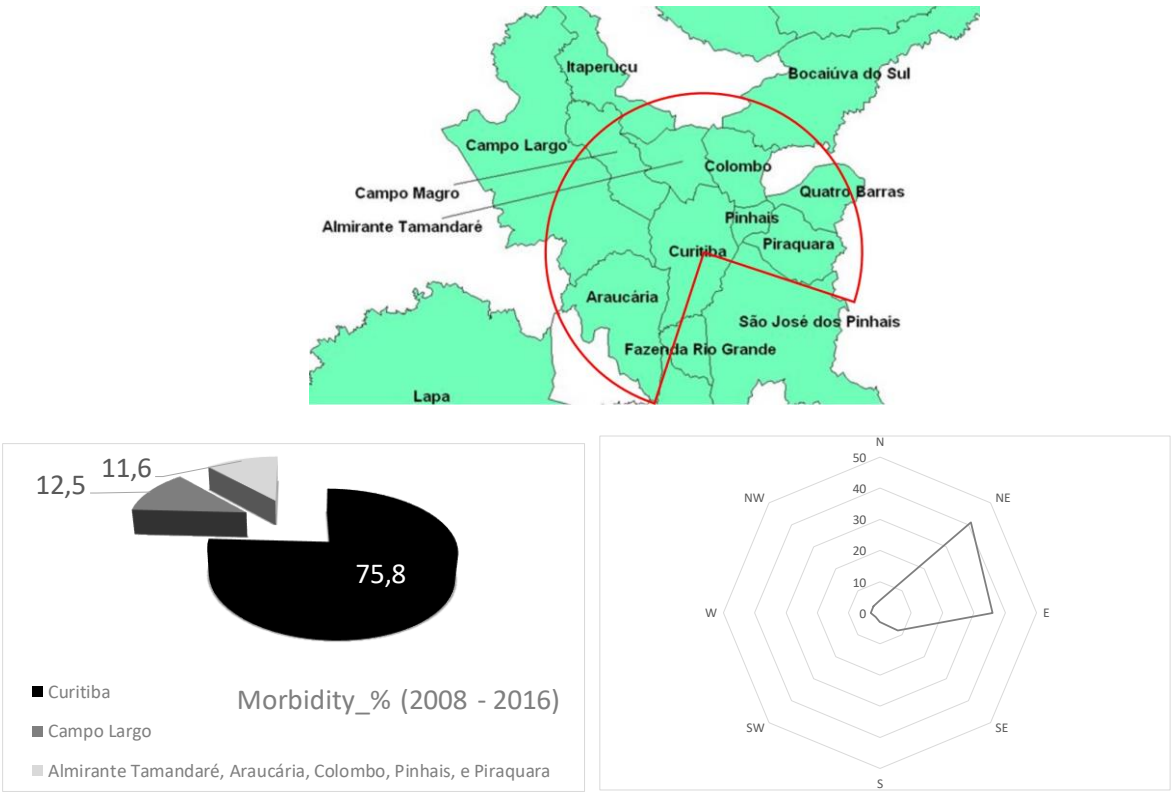
Fonte: Autor, IAP, INMET, DATASUS.

Prevalência Sazonal da Morbidade

A Figura 2 ilustra a Região Metropolitana de Curitiba e uma semicircunferência em vermelho com a projeção do centroide de Curitiba e os limites de direção delimitados pelas áreas abrangidas das seis cidades circunvizinhas (topo). Abaixo e à esquerda um gráfico com o percentual de morbidade das sete cidades. Percebe-se que Curitiba é responsável por aproximadamente 3/4 da morbidade entre as sete cidades escolhidas e que em média registra 40 casos diários de internações hospitalares por doenças respiratórias entre os anos de 2008 e 2016. Campo Largo é a segunda cidade das sete e responsável por cerca de 1/8 da morbidade. As cinco outras cidades circunvizinhas juntas somam aproximadamente outro 1/8. Abaixo e à direita é mostrado um gráfico com a predominância da direção do vento. Percebe-

se maior incidência de ventos vindo da direção nordeste e leste, e, portanto, a escolha das setes cidades contempla os principais padrões de circulação do vento na região. Além disso, as direções predominantes do gráfico instigam maiores análises para verificar se o alto índice de morbidade em Campo Largo está associado à pluma de poluentes gerada na própria cidade ou à pluma oriunda de Curitiba é deslocada para oeste e sudeste onde se localiza Campo Largo.

Figura 2. Área Metropolitana de Curitiba (topo); percentual de morbidade das sete cidades (inferior à esquerda); predomínio da direção do vento (inferior à direita).



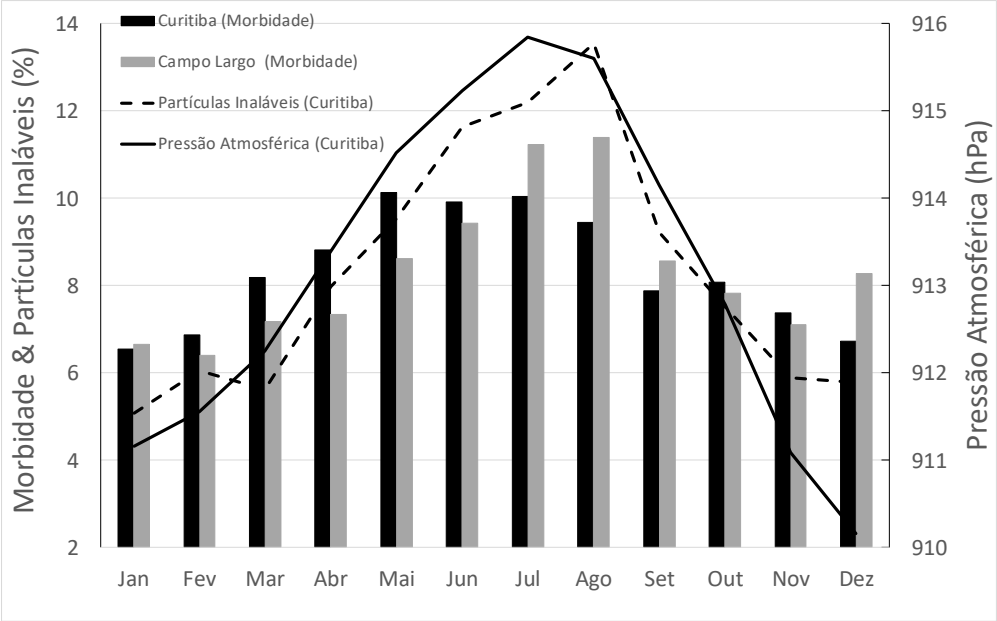
Fonte: Autor, INMET, DATASUS.

A Figura 3 ilustra a variação sazonal dos percentuais de morbidade em Curitiba e Campo Largo, da predominância do percentual de partículas inaláveis medidas na estação Ouvidor Pardinho no centro de Curitiba, além da predominância dos valores médios de pressão atmosférica (hPA) medidos em Curitiba.

No Brasil, o verão termina em março e a partir do início do outono até final de agosto existe uma progressiva diminuição da temperatura que implica no aumento gradativo da densidade do ar. Este fenômeno é claramente observado pelos registros do aumento da pressão atmosférica na região. Simultaneamente neste período existe uma diminuição da velocidade do vento com dano à dispersão dos contaminantes do ar. O ar mais denso aprisiona

as plumas de poluentes nas áreas urbanas e é nesse período de outono e inverno, portanto, que ocorre um aumento nos índices de morbidade nas regiões metropolitanas.

Figura 3. Percentual de morbidade de Curitiba e Campo Largo e predominância de pressão atmosférica e partículas inaláveis.



Fonte: Autor, IAP, INMET, DATASUS.

Previsão da Morbidade

Foram desenvolvidos dois classificadores (modelos 1 e 2) para prever três níveis diferentes de morbidade por quatro períodos distintos (próximo, segundo, terceiro e quarto meses). Os níveis de morbidade foram calculados considerando-se os tercis (*baixo*, *médio* e *alto*). Os modelos preditivos foram construídos usando regras de classificação. No primeiro passo, o algoritmo varre todo o conjunto de dados e encontra um conjunto de regras de forte suporte (união entre variáveis de entrada e saída). Na segunda etapa (validação do modelo), o algoritmo aplica o conjunto de regras para prever o nível de morbidade. O resultado da validação do classificador é geralmente apresentado em uma matriz de confusão onde a diagonal principal mostra a classificação correta e a classificação incorreta é mostrado fora dessa diagonal. A acurácia do modelo (taxa de classificação correta) é a soma do amostras classificadas corretamente dividido pelo número total de amostras na validação. Os dois classificadores correspondem à combinação de duas abordagens estatísticas (modelo 1, análise fatorial e correlação linear) e duas abordagens de agrupamento de dados (modelo 2, dendrograma e método *k-means*).

A Tabela 5 mostra a matriz de confusão para a previsão do nível de morbidade no próximo mês (T1), onde se observa apenas um erro de classificação. A taxa de classificação

correta (TCC), neste caso, é a soma da diagonal principal (30 + 34 + 38 = 102) dividido pelo total número de amostras (103) e multiplicado por 100 (TCC = 99,03%).

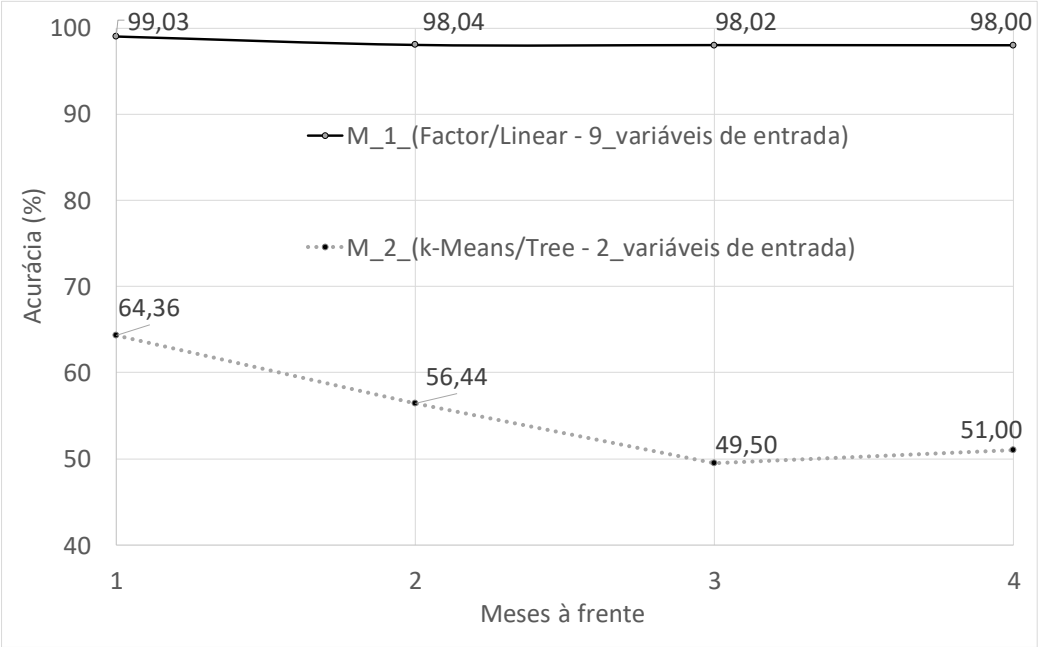
Tabela 5. Matriz de confusão para a previsão do nível de morbidade no próximo mês (T1).

	M_S_CWB_DR_T1_<_1100	1100_<_M_S_CWB_DR_T1_<_1250	M_S_CWB_DR_T1_>_1250
M_S_CWB_DR_T1_<_1100	30	0	0
1100_<_M_S_CWB_DR_T1_<_1250	0	34	0
M_S_CWB_DR_T1_>_1250	0	1	38

Fonte: Autor.

A Figura 4 mostra a acurácia dos dois modelos para os horizontes temporais de 1, 2, 3 e 4 meses à frente. A acurácia ou TCC é excelente (igual ou superior a 98%) para o modelo 1 ainda que para o quarto à frente (T4). O Modelo 1 considera nove variáveis explicativas selecionadas da análise fatorial e correlação linear. O modelo 2 tem precisão menor que 65%, e seu desempenho diminui até o quarto mês à frente. O modelo 1 poderia ser utilizado para ajudar as autoridades alertar sobre potenciais surtos e promover a mitigação das doenças respiratórias em centros urbanos. Os modelos de previsão de morbidade nas outras cidades também apresentam excelente desempenho (Cardoso & Souza, 2017).

Figura 4. Acurácia dos dois modelos para os horizontes temporais de 1, 2, 3 e 4 meses à frente.



Fonte: Autor.

São apresentadas algumas regras extraídas dos classificadores, tais como segue:

Regra 3:

SE Janeiro (A)

ENTÃO → Classe = M_S_C_CB_DR_T1 _ <_ 1100 (B)

(8,738% 100 % 9 9 8,738%)

Se o mês atual é janeiro (A), ENTÃO, morbidade por doença respiratória em Curitiba é menor que 1100 pacientes no próximo mês

(8,738% 100,000% 9 9 8,738%)

Das nove vezes que (A) ocorre no banco de dados, (B) também ocorre; portanto, essa regra tem 100% de confiança (9/9 × 100). Ao longo do período analisado (2008 a 2016), o mesmo padrão é observado para fevereiro.

Analogamente, as seguintes regras descrevem surtos para maio e julho com mais de 1250 pacientes internados para esses meses.

Regra 4:

SE Abril (A)

ENTÃO → Classe = M_S_C_CB_DR_T1 _> _ 1250 (B)

(8,738% 100% 9 9 8,738%)

Regra 5:

SE Junho (A)

ENTÃO → Classe = M_S_C_CB_DR_T1 _> _ 1250 (B)

(8,738% 100,000% 9 9 8,738%)

Cabe ressaltar que as regras 4 e 5 elucidam um grave problema para estes dois meses do ano e tais regras, portanto, poderiam ser usadas imediatamente pelo governo como uma ferramenta para ajudar na redução das hospitalizações por doenças respiratórias.

CONCLUSÕES

Este artigo descreveu uma metodologia para consistência e preparação de dados, bem como uma análise preliminar de alguns padrões sazonais de variáveis atmosféricas e os efeitos na morbidade por doenças respiratórias. Os padrões sazonais identificados mostram um

aumento na poluição do ar após o final do verão, com consequente aumento dos níveis de morbidade até os meses de julho/agosto. Este fenômeno também é explicado por variações na pressão atmosférica e na densidade do ar. Essa variação orienta a dispersão de poluentes na atmosfera e foi incluído na criação de modelos para prever níveis de morbidade nas sete cidades, com até 4 meses de antecedência.

Os padrões de direção do vento também mostram a importância de estudos aprofundados sobre circulação do ar na Região Metropolitana de Curitiba para a descoberta de fontes da poluição do ar e das relações com as regiões afetadas pelas plumas de poluição. No entanto, mais estações meteorológicas devem ser consideradas nesta análise, a fim de explicar o comportamento espaço-temporal do vento na região.

As cidades analisadas apresentam padrões de morbidade peculiares e exigem investigações mais detalhadas das variáveis envolvidas nos fenômenos urbanos. As taxonomias qualidade do ar, doenças respiratórias e climatologia foram considerados na coleta de dados. No entanto, devido à complexidade dos fenômenos estudados, seria benéfico expandir o número de taxonomias envolvidas, incluindo outras tais como o uso e ocupação do solo urbano, classificação do grau de poluição industrial, principais vias de tráfego de veículos poluentes, proximidades de parques industriais, estudos da camada limite, entre outras.

Os modelos preditivos apresentam excelente acurácia na validação e poderiam ser usados pelo governo para estabelecer ações e intervenções a fim de minimizar os efeitos nocivos à saúde causados pela poluição atmosférica, bem como propostas de melhoraria da qualidade do ar e do bem-estar da população. Espera-se que estes resultados possam contribuir para o planejamento de saúde urbana e para a melhoria de políticas públicas. Além disso, a metodologia apresentada aqui poderia ser estendida para megacidades que estão vulneráveis às mudanças climáticas causadas pelas emissões de CO2 dos combustíveis fósseis. Muitos pesquisadores enfatizam a necessidade de entender o ciclo do carbono e enfatizam a urgência da quantificação das emissões com resoluções temporais e espaciais adequadas (Newman et al., 2013; Gurney et al., 2012; Kort et al., 2012; Gurney et al., 2009).

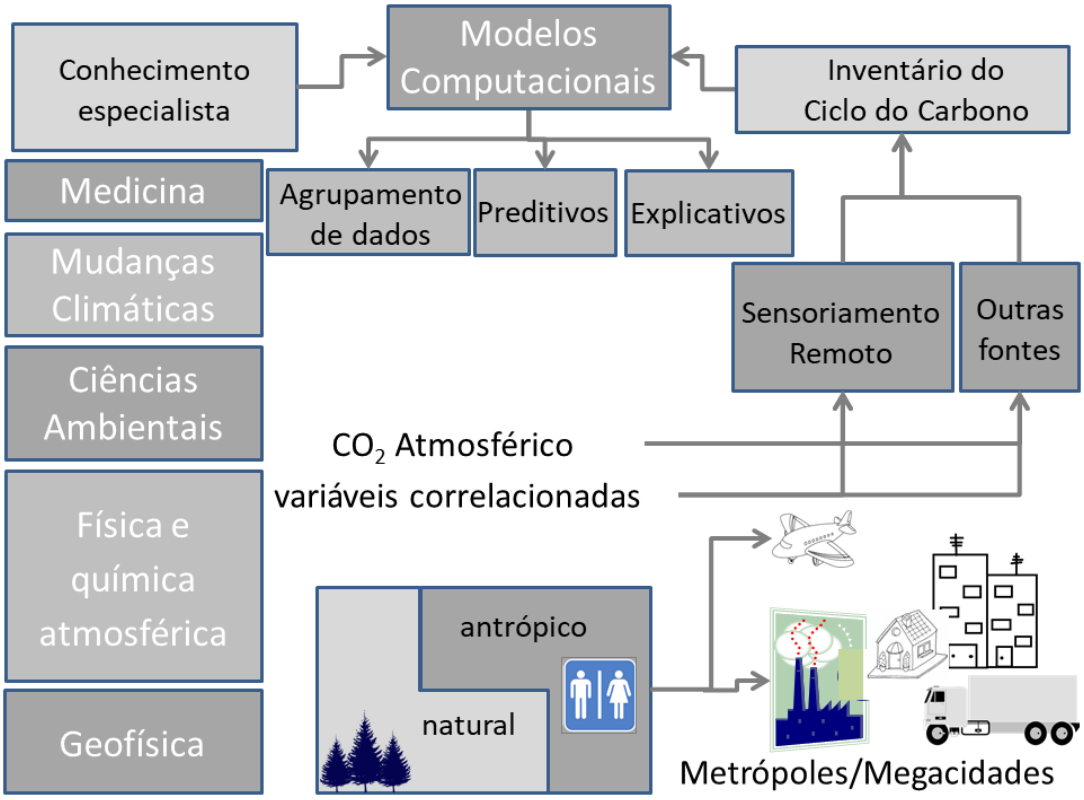
A complexidade do monitoramento e entendimento da variabilidade espaço- temporal das fontes de emissão é discutida na literatura atual, e a manutenção de uma constelação de satélites para monitorar o CO2 atmosférico é sugerido (Schwandner et al., 2014). A Figura 5 mostra um modelo de monitoramento de CO2 e variáveis correlacionáveis em megacidades, inclui o suporte do conhecimento especialista de diferentes áreas de pesquisa e modelagem de dados. Algumas aplicações potenciais de técnicas de mineração de dados em dados de sensoriamento remoto e em outras fontes podem também ser considerado em pesquisas futuras.

RECOMENDAÇÕES PARA FUTUROS ESTUDOS

Esta pesquisa considera o desenvolvimento de metodologias distintas em mineração de dados e a integração do conhecimento de especialistas e dados de diferentes áreas do ciclo

do carbono, tais como de saúde coletiva, medicina, mudanças climáticas, gestão ambiental, física e química atmosférica, geofísica, entre outras. Patton et al. (2015) ressaltam a importância da construção de modelos transferíveis que sejam aplicáveis a outros locais (transferência) e modelos gerais aplicáveis a áreas menores (generalização). Além disso, Liu et al. (1994) encorajam estudos mais aprofundados da circulação de poluentes e dos efeitos epidemiológicos dos padrões de vento. A integração do conhecimento especialista e dos inventários de dados seria útil para explicar os fenômenos complexos que envolvem as emissões de CO2 nas megacidades, como por exemplo São Paulo, Rio de Janeiro ou Bogotá. Modelos computacionais que incorporem diferentes perspectivas, como agrupamento de dados, modelos preditivos e modelos explicativos devem ser desenvolvidos. A integração dos dados de satélite e de superfície de várias cidades em um modelo nacional poderia ser uma estratégia cientificamente interessante para novas descobertas de padrões associados aos fenômenos descritos neste artigo.

Figura 5. Monitoramento de CO2 atmosférico e modelo de conhecimento.



Fonte: Autor.

REFERÊNCIAS

AGRAWAL, Rakesh, & Ramakrishnan Srikant. "Fast algorithms for mining association rules." In Proc. 20th int. conf. very large data bases, VLDB, vol. 1215, pp. 487-499. 1994.

AGRAWAL, Rakesh, Tomasz Imieliński, & Arun Swami. "Mining association rules between sets of items in large databases." In ACM sigmod record, vol. 22, no. 2, pp. 207-216. 1993.

BAKONYI, Sonia Maria Cipriano, Inês Moresco Danni-Oliveira, Lourdes Conceição Martins, & Alfésio Luís Ferreira Braga. "Poluição atmosférica e doenças respiratórias em crianças na cidade de Curitiba, PR." Revista de Saúde Pública 38: 695-700. 2004.

BECERRA, Janeth Mosquera, Rodrigo S. Reis, Lawrence D. Frank, Farah A. Ramirez-Marrero, Benjamin Welle, Eugenio Arriaga Cordero, Fabian Mendez Paz et al. "Transport and health: a look at three Latin American cities." Cadernos de Saúde Pública 29: 654-666. 2013.

BRAVERMAN, Amy. A Statistical Perspective on Climate Informatics. NASA report. 2012.

BROWNSON, Ross C., Rodrigo S. Reis, Peg Allen, Kathleen Duggan, Robert Fields, Katherine A. Stamatakis, & Paul C. Erwin. "Understanding administrative evidence-based practices: findings from a survey of local health department leaders." American journal of preventive medicine 46, no. 1: 49-57. 2014.

CARDOSO, Matheus Bittencourt & Souza, Fabio Teodoro. "Prediction of Hospitalizations caused by respiratory diseases by using data mining techniques: some applications in Curitiba, Brazil and the Metropolitan Area." WIT Transactions on Ecology and the Environment 211: 231-241. 2017.

D´ANGIOLA, Ariela, Laura E. Dawidowski, Darío R. Gómez, & Mauricio Osses. "On-road traffic emissions in a megacity." Atmospheric Environment 44, no. 4: 483-493. 2010.

DAVIS, Craig A., & J. P. Bardet. "Lifelines in megacities." In Geotechnics and earthquake geotechnics towards global sustainability, pp. 37-58. Springer, Dordrecht, 2011.

DE MOURA, Maria Augusta Costa. "A Urbanização em Campina Grande e suas relações com a incidência de doenças respiratórias no município e o clima local." UFCG. Programa de Pós-Graduação em Recursos Naturais. 2009.

DECKER, Ethan H., Scott Elliott, Felisa A. Smith, Donald R. Blake, & F. Sherwood Rowland. "Energy and material flow through the urban ecosystem." Annual review of energy and the environment 25. 2000.

DUARTE, Fábio, Tatiana Gadda, Carlos Augusto Moreno Luna, & Fábio Teodoro Souza. "What to expect from the future leaders of bogotá and curitiba in terms of public transport: Opinions and practices among university students." Transportation research part F: traffic psychology and behaviour 38: 7-21. 2016.

DUCHIADE, Milena P. Poluição do ar e doenças respiratórias: uma revisão. Cadernos de Saúde Pública, v. 8, n. 3, p. 311-30, 1992.

DUREN, Riley M., & Charles E. Miller. "Measuring the carbon emissions of megacities." Nature Climate Change 2, no. 8: 560. 2012.

FAGHMOUS, James H., & Vipin Kumar. "Spatio-temporal data mining for climate data: Advances, challenges, and opportunities." In Data mining and knowledge discovery for big data, pp. 83-116. Springer, Berlin, Heidelberg, 2014.

GRIMM, Nancy B., Stanley H. Faeth, Nancy E. Golubiewski, Charles L. Redman, Jianguo Wu, Xuemei Bai, & John M. Briggs. "Global change and the ecology of cities." science 319, no. 5864: 756-760. 2008.

GUAN, Dabo, Zhu Liu, Yong Geng, Sören Lindner, & Klaus Hubacek. "The gigatonne gap in China's carbon dioxide inventories." Nature Climate Change 2, no. 9: 672. 2012.

GULLIVER, Gulliver, John, & David J. Briggs. "Time–space modeling of journey-time exposure to traffic-related air pollution using GIS." Environmental research 97, no. 1: 10-25. 2005.

GURJAR, Bhola R., Luisa T. Molina, & C. Shekhar P. Ojha. Air pollution: health and environmental impacts. CRC press, 2010.

GURNEY, Kevin R., Daniel L. Mendoza, Yuyu Zhou, Marc L. Fischer, Chris C. Miller, Sarath Geethakumar, & Stephane de la Rue du Can. "High resolution fossil fuel combustion CO2 emission fluxes for the United States." Environmental science & technology 43, no. 14: 5535-5541. 2009.

GURNEY, Kevin R., Igor Razlivanov, Yang Song, Yuyu Zhou, Bedrich Benes, & Michel Abdul-Massih. "Quantification of fossil fuel CO2 emissions on the building/street scale for a large US city." Environmental science & technology 46, no. 21 (): 12194-12202. 2012.

GUTTIKUNDA, Sarath K., & Giuseppe Calori. "A GIS based emissions inventory at 1 km× 1 km spatial resolution for air pollution analysis in Delhi, India." Atmospheric Environment 67: 101-111. 2013.

HINO, Adriano AF, Rodrigo S. Reis, Olga L. Sarmiento, Diana C. Parra, & Ross C. Brownson. "Built environment and physical activity for transportation in adults from Curitiba, Brazil." Journal of urban health 91, no. 3: 446-462. 2014.

IAP (Instituto Ambiental do Paraná), Relatório de Monitoramento da Qualidade do Ar (em tempo real), 2014: disponível em: http://www.iap.pr.gov.br/modules/conteudo/conteudo.php?conteudo=1076

KAN, Haidong, Renjie Chen, & Shilu Tong. "Ambient air pollution, climate change, and population health in China." Environment international 42: 10-19. 2012.

KENNEDY, Christopher, John Cuddihy, & Joshua Engel-Yan. "The changing metabolism of cities." Journal of industrial ecology 11, no. 2: 43-59. 2007.

KORT, Eric A., Christian Frankenberg, Charles E. Miller, & Tom Oda. "Space-based observations of megacity carbon dioxide." Geophysical Research Letters 39, no. 17. 2012.

LIU, Bing, Wynne Hsu, Shu Chen, & Yiming Ma. "Analyzing the subjective interestingness of association rules." IEEE Intelligent Systems and their Applications 15, no. 5: 47-55. 2000.

LIU, Hai-Ying, Erik Skjetne, & Mike Kobernus. "Mobile phone tracking: in support of modelling traffic-related air pollution contribution to individual exposure and its implications for public health impact assessment." Environmental Health 12, no. 1: 93. 2013.

LIU, L-J. Sally, Melvin P. Olson, George A. Allen, Petros Koutrakis, William F. McDonnell, & Timothy R. Gerrity. "Evaluation of the Harvard ozone passive sampler on human subjects indoors." Environmental science & technology 28, no. 5: 915-923. 1994.

LIU, Ma, Bing Liu Wynne Hsu Yiming, & Bing Liu. "Integrating classification and association rule mining." In Proc. of the 4th Int. Conf. on knowledge discovery and data mining. 1998.

LU, Heli, & Guifang Liu. "Spatial effects of carbon dioxide emissions from residential energy consumption: A county-level study using enhanced nocturnal lighting." Applied energy 131: 297-306. 2014.

MARTINS, Lourdes Conceição, Maria do Rosário Dias de Latorre, Maria Regina Alves Cardoso, Fábio Luiz Teixeira Gonçalves, Paulo Hilário Nascimento Saldiva, & Alfésio Luís Ferreira Braga. "Poluição atmosférica e atendimentos por pneumonia e gripe em São Paulo, Brasil." Revista de Saúde Pública 36: 88-94. 2002.

MERBITZ, H. M. Buttstädt, S. Michael, W. Dott, & C. Schneider. "GIS-based identification of spatial variables enhancing heat and poor air quality in urban areas." Applied Geography 33: 94-106. 2012.

MORAGUES, Amparo, & Teresa Alcaide. "The use of a geographical information system to assess the effect of traffic pollution." Science of the total environment 189: 267-273. 1996.

MORIK, Katharina, Kanishka Bhaduri, & Hillol Kargupta. "Introduction to data mining for sustainability." Data Mining and Knowledge Discovery 24, no. 2: 311-324. 2012.

NEWMAN, S., S. Jeong, M. L. Fischer, X. Xu, C. L. Haman, B. Lefer, S. Alvarez et al. "Diurnal tracking of anthropogenic CO 2 emissions in the Los Angeles basin megacity during spring 2010." Atmospheric Chemistry and Physics 13, no. 8: 4359-4372. 2013.

PATTON, Patton, Allison P., Wig Zamore, Elena N. Naumova, Jonathan I. Levy, Doug Brugge, & John L. Durant. "Transferability and generalizability of regression models of ultrafine particles in urban neighborhoods in the Boston area." Environmental science & technology 49, no. 10: 6051-6060. 2015.

PINZÓN, Daniel Felipe Del Busto, & Fabio Teodoro de Souza. "A data based model as a metropolitan management tool: The Bogotá-Sabana region case study in Colombia." Land Use Policy 54: 253-263. 2016.

PYLE, Dorian. Data preparation for data mining. Morgan Kaufmann, 1999.

SCHWANDNER, Florian M., Charles E. Miller, Riley M. Duren, Vijay Natraj, Annmarie Eldering, Michael R. Gunson, & David Crisp. "Strategies for satellite-based monitoring of CO2 from distributed area and point sources." In EGU General Assembly Conference Abstracts, vol. 16. 2014.

SCOGGINS, Amanda, Tord Kjellstrom, Gavin Fisher, Jennie Connor, & Neil Gimson. "Spatial analysis of annual air pollution exposure and mortality." Science of the Total Environment 321, no. 1-3: 71-85. 2004.

SHAH, Anoop SV, Jeremy P. Langrish, Harish Nair, David A. McAllister, Amanda L. Hunter, Ken Donaldson, David E. Newby, & Nicholas L. Mills. "Global association of air pollution and heart failure: a systematic review and meta-analysis." The Lancet 382, no. 9897: 1039-1048. 2013.

SILVA, Sam J., Avelino F. Arellano, & Helen M. Worden. "Toward anthropogenic combustion emission constraints from space-based analysis of urban CO2/CO sensitivity." Geophysical Research Letters 40, no. 18: 4971-4976. 2013.

SOUZA, Fabio Teodoro. "Morbidity Forecast in Cities: A Study of Urban Air Pollution and Respiratory Diseases in the Metropolitan Region of Curitiba, Brazil." Journal of Urban Health: 1-14. 2018.

SOUZA, Fábio Teodoro., & Nelson FF Ebecken. "A data based model to predict landslide induced by rainfall in Rio de Janeiro city." Geotec. & Geolog Eng., 30, no. 1: 85-94. 2012.

VON DER WEINDEN-REINMÜLLER, S-L., F. Drewnick, M. Crippa, A. S. H. Prévôt, Frédérik Meleux, U. Baltensperger, M. Beekmann, & S. Borrmann. "Application of mobile aerosol and trace gas measurements for the investigation of megacity air pollution emissions: the Paris metropolitan area." Atmospheric Measurement Techniques 7, no. 1: 279-299. 2014.

WANG, Xinhao. "Integrating GIS, simulation models, and visualization in traffic impact analysis." Computers, Environment and Urban Systems 29, no. 4: 471-496. 2005.

WENTZ, Elizabeth A., Sharolyn Anderson, Michail Fragkias, Maik Netzband, Victor Mesev, Soe W. Myint, Dale Quattrochi, Atiqur Rahman, & Karen C. Seto. "Supporting global environmental change research: A review of trends and knowledge gaps in urban remote sensing." Remote Sensing 6, no. 5: 3879-3905. 2014.

ZHANG, Qingyu, Yumei Wei, Weili Tian, & Kemin Yang. "GIS-based emission inventories of urban scale: A case study of Hangzhou, China." Atmospheric Environment 42, no. 20: 5150-5165. 2008.