

Modelo fuzzy estimando tempo de internação por doenças cardiovasculares

Fuzzy model approach for estimating time of hospitalization due to cardiovascular diseases

Karine Mayara Vieira Coutinho ¹
Paloma Maria Silva Rocha Rizol ¹
Luiz Fernando Costa Nascimento ²
Andréa Paula Peneluppi de Medeiros ²

Abstract A fuzzy linguistic model based on the Mamdani method with input variables, particulate matter, sulfur dioxide, temperature and wind obtained from CETESB with two membership functions each was built to predict the average hospitalization time due to cardiovascular diseases related to exposure to air pollutants in São José dos Campos in the State of São Paulo in 2009. The output variable is the average length of hospitalization obtained from DATASUS with six membership functions. The average time given by the model was compared to actual data using lags of 0 to 4 days. This model was built using the Matlab v. 7.5 fuzzy toolbox. Its accuracy was assessed with the ROC curve. Hospitalizations with a mean time of 7.9 days (SD = 4.9) were recorded in 1119 cases. The data provided revealed a significant correlation with the actual data according to the lags of 0 to 4 days. The pollutant that showed the greatest accuracy was sulfur dioxide. This model can be used as the basis of a specialized system to assist the city health authority in assessing the risk of hospitalizations due to air pollutants.

Key words Fuzzy logic, Air pollutants, Cardiovascular diseases, Sulfur dioxide, Particulate matter

Resumo Para prever o tempo médio de internações por doenças cardiovasculares relacionadas à exposição de poluentes do ar em São José dos Campos (SP), em 2009, foi construído um modelo linguístico fuzzy, baseado no método de Mamdani, com variáveis de entrada: material particulado, dióxido de enxofre, temperatura e vento, obtidos da CETESB, com duas funções de pertinência cada. A variável de saída é o tempo médio de internações, obtido do Datasus, com seis funções de pertinência. O tempo médio fornecido pelo modelo foi comparado aos dados reais segundo defasagens de 0 a 4 dias. Este modelo foi construído no toolbox fuzzy do Matlab v. 7.5. Sua acurácia foi avaliada pela curva ROC. Foram registradas 1119 internações com o tempo médio de 7,9 dias ($dp = 4,9$). Os dados fornecidos mostraram significativa correlação com os dados reais, segundo as defasagens de 0 a 4 dias. O poluente que mostrou melhor acurácia foi o dióxido de enxofre. Este modelo pode ser utilizado como base de sistema especialista, que pode auxiliar o gestor municipal na avaliação do risco de internações em função dos poluentes do ar.

Palavras-chave Lógica fuzzy, Poluentes do ar, Doenças cardiovasculares, Dióxido de enxofre, Material particulado

¹ Departamento de Engenharia Elétrica, Faculdade de Engenharia de Guaratinguetá, Universidade Estadual Paulista. Av. Ariberto Pereira Cunha 333, Pedregulho. 12516-410 Guaratinguetá SP Brasil. luiz.nascimento@pq.cnpq.br

² Departamento de Medicina, Universidade de Taubaté.

Doenças cardiovasculares e lógica fuzzy

Diversos estudos epidemiológicos têm apontado a poluição do ar como um dos fatores que oferece maior impacto na saúde humana, associado principalmente a doenças dos sistemas respiratório¹⁻⁶ e circulatório⁶⁻¹⁰. Em cidades que constituem importantes polos industriais, como é o caso de São José dos Campos, devido ao intenso tráfego de automóveis e caminhões, alta densidade populacional e importante desenvolvimento industrial, a população enfrenta sérios problemas ambientais como o elevado acúmulo de poluentes atmosféricos.

Segundo o Datasus – Departamento de Informática do Sistema Único de Saúde (SUS), em 2012 foi registrado 1,1 milhão de internações por doenças cardiovasculares no Brasil, sendo 600 mil no Estado de São Paulo, gerando despesa de 2,4 bilhões aos cofres públicos¹¹.

Os poluentes estudados e mais comumente associados a efeitos deletérios sobre a saúde humana são o material particulado (PM_{10}), dióxido de enxofre (SO_2), óxido de nitrogênio (NO e NO_2) e ozônio (O_3)¹. O PM_{10} é uma mistura de partículas sólidas e líquidas, com diâmetro inferior a $10\mu m$, suspensas no ar e produzidas pela queima de combustíveis e biomassa; o SO_2 é gerado a partir da combustão de elementos fósseis e o O_3 é formado por reação entre a radiação ultravioleta, os óxidos de nitrogênio e hidrocarbonetos emitidos por veículos¹².

Estudos mostram que indivíduos expostos a altos índices de poluição apresentam aumento da pressão arterial, variabilidade da frequência cardíaca, o que proporciona maior risco de arritmia e morte súbita, além de infarto e acidente vascular encefálico e uma alta incidência de inflamação brônquica, acompanhada de declínio acentuado da função pulmonar^{2,4,7,10}.

Acredita-se que o material particulado (PM_{10}) e o dióxido de enxofre (SO_2) estão intimamente relacionados a internações e mortes por insuficiência cardíaca, riscos que se mostram mais fortes no dia em que ocorre a exposição aos poluentes⁷.

Atualmente, em muitos estudos nesta área, são utilizadas técnicas estatísticas como a regressão logística e a regressão de Poisson (Modelos Lineares Generalizados-MLG ou Modelos Aditivos Generalizados-MAG)⁵. Estes modelos são capazes de estimar a chance ou o risco de internação ou óbito por doenças respiratórias e cardiovasculares. No entanto, modelos *fuzzy* surgem como uma nova opção de ferramenta, devido a sua capacidade de lidar com a imprecisão e in-

certeza das informações¹³, onde os valores podem ser classificados como parcialmente verdadeiros, possui facilidade de compreensão e baixo custo computacional. Esta abordagem tem sido utilizada em diversas áreas como em controle de processos industriais¹⁴⁻¹⁶ e medicina^{2,4,17,18}.

O conceito de Lógica *fuzzy* foi proposto por Lotfi Zadeh e sua teoria é baseada em conceitos da lógica clássica, porém com a definição de graus de pertinência^{13,19}.

Na teoria clássica, os conjuntos são denominados *crisp*, onde um elemento de um universo de discurso pode ou não pertencer a um determinado conjunto. Na teoria dos conjuntos *fuzzy*, como existe um grau de pertinência de cada elemento a um determinado conjunto, não existe uma fronteira a partir da qual é possível definir se um elemento pertence ao conjunto em questão, sendo que se tornou importante a abordagem matemática a ser aplicada na área de saúde, provavelmente devido à sua capacidade de lidar com termos linguísticos.

Nesta área, a lógica *fuzzy* tem sido desenvolvida na modelagem do processo de diagnósticos, padrões *fuzzy* de reconhecimento e processamento de imagem para análise de fotos de raio-x, em epidemiologia e saúde pública^{2,4,17,18,20}.

O objetivo deste trabalho é construir um modelo computacional utilizando as propriedades da lógica *fuzzy* para estimar o tempo médio de internações por doenças cardiovasculares, com base nas concentrações de poluentes do ar na cidade de São José dos Campos, Brasil.

Métodos

Trata-se de um modelo computacional utilizando ferramentas da lógica *fuzzy* para estimar o tempo médio de internação por doenças cardiovasculares, segundo concentrações de PM_{10} (material particulado), SO_2 (dióxido de enxofre) e valores de temperatura mínima e de velocidade do vento, obtidos da Companhia de Tecnologia de Saneamento Ambiental (CETESB), que conta com uma estação medidora no município de São José dos Campos (SP)²¹.

O modelo foi desenvolvido com base em dados reais das internações por doenças cardiovasculares em indivíduos, de todas as idades, residentes na cidade de São José dos Campos, no período de 01/01/2009 a 31/12/2009. Esses dados foram obtidos do Departamento de Informações e Informática do Sistema Único de Saúde (Datasus), referentes aos diagnósticos do capítulo IX.

São José dos Campos é uma cidade de médio porte do interior do Estado de São Paulo que possui importante parque industrial. Localiza-se no Alto Vale do Paraíba, a 660m acima do nível do mar, tem clima úmido e tropical de altitude, localizando-se entre São Paulo e Rio de Janeiro – as duas maiores cidades do Brasil; é cortada pela Via Dutra, rodovia mais importante do país e que tem tráfego intenso de ônibus e caminhões durante todo o ano, cerca de 130 mil veículos por dia. Sua população é estimada em torno de 700 mil habitantes²².

Este modelo é baseado na *fuzzificação* das variáveis: material particulado (PM_{10}), dióxido de enxofre (SO_2), temperatura mínima e vento.

O modelo foi desenvolvido com o auxílio de um especialista que elaborou duas funções de pertinência, Aceitável e Inaceitável, para as variáveis de entrada: material particulado (PM_{10}) e dióxido de enxofre (SO_2), Boa e Ruim para a variável temperatura mínima (T_{MIN}) e Fraco e Forte para a variável de entrada velocidade do vento. A saída foi o tempo médio de internações, com seis funções de pertinência: muito baixo, baixo, médio, médio alto, alto e muito alto. Os universos de discurso em que os índices de cada variável eram considerados baixos ou altos foram estipulados com base no quão nocivo à saúde é o elemento em questão.

Ao se combinarem todas as possíveis entradas, cada uma com duas funções de pertinência, foi possível a construção de 16 regras ($2 \times 2 \times 2 \times 2$).

O procedimento do modelo linguístico *fuzzy* consiste no cálculo do grau de pertinência das variáveis: PM_{10} , SO_2 , T_{MIN} e Vento. Em seguida, é realizado o processo de inferência *fuzzy* proposto por Mamdani; o número de internações é estimado pela técnica de defuzzificação baseada no método centro de área, conforme descrito em artigo anterior¹⁷.

Através da rotina utilizando o programa MATLAB®, é possível gerar uma saída numérica para o modelo *fuzzy*, portanto, obtém-se o tempo médio de internação a partir dos valores de entrada de cada variável. A partir desta resposta, compara-se o tempo médio de internações obtido aos dados de internações reais, utilizando até dois dias de defasagem.

Para a avaliação dos resultados foi feita a correlação de Pearson e a acurácia foi determinada utilizando-se a curva ROC.

Por se tratar de dados disponíveis na rede e não identificados, foi dispensada a submissão a Comitê de Ética em Pesquisa.

Resultados

No período foram registradas 1119 internações na cidade de São José dos Campos (SP), com diagnóstico de doenças cardiovasculares.

As variáveis de entrada, material particulado (PM_{10}), dióxido de enxofre (SO_2), temperatura mínima (T_{MIN}) e Vento, são apresentadas na Figura 1(a-d). A variável de saída: tempo médio de internação por doenças cardiovasculares, com suas respectivas funções de pertinência, e apresentada na Figura 2.

Os valores médios, desvios-padrão, valores mínimo e máximo das variáveis material particulado (PM_{10}), dióxido de enxofre (SO_2), temperatura mínima, vento e tempo médio de internações estão apresentados na Tabela 1.

Os resultados fornecidos pela curva ROC são mostrados na Tabela 2. O melhor resultado foi para a defasagem do dia zero, sendo o PM_{10} e SO_2 aqueles que mostraram melhores desempenhos evidentes pela curva ROC, com valores estatisticamente significativos.

A correlação de Pearson entre os resultados fornecidos pelo modelo *fuzzy* e os dados reais foi de 0,22 ($p < 0,01$) para o lag 0, 0,15 para o lag 1 e 0,23 para o lag 2, mesmo não sendo valores elevados, são significativos.

Discussão

Até onde se tem conhecimento, este é o primeiro estudo realizado no Brasil envolvendo o efeito da exposição a poluentes do ar sobre o tempo médio de internação por doenças cardiovasculares, usando ferramenta de lógica *fuzzy*.

Este estudo trata da construção de um modelo computacional, com abordagem *fuzzy*, para estimar o tempo médio de internações por doenças cardiovasculares relacionadas aos poluentes do ar em uma cidade industrial de médio porte, mostrando boa acurácia para a predição do tempo médio de internação quando a exposição ocorreu no mesmo dia e até dois dias depois.

Estudos sobre os impactos da poluição do ar na saúde humana geralmente se utilizam de ferramentas, como regressão logística e de Poisson, onde geralmente não é considerado o tempo de internação dos pacientes, entretanto, este constitui uma variável de extrema importância visto que representa o efetivo custo para a sociedade. Em relação a estes modelos, a abordagem *fuzzy* possui a vantagem de lidar com a incerteza das

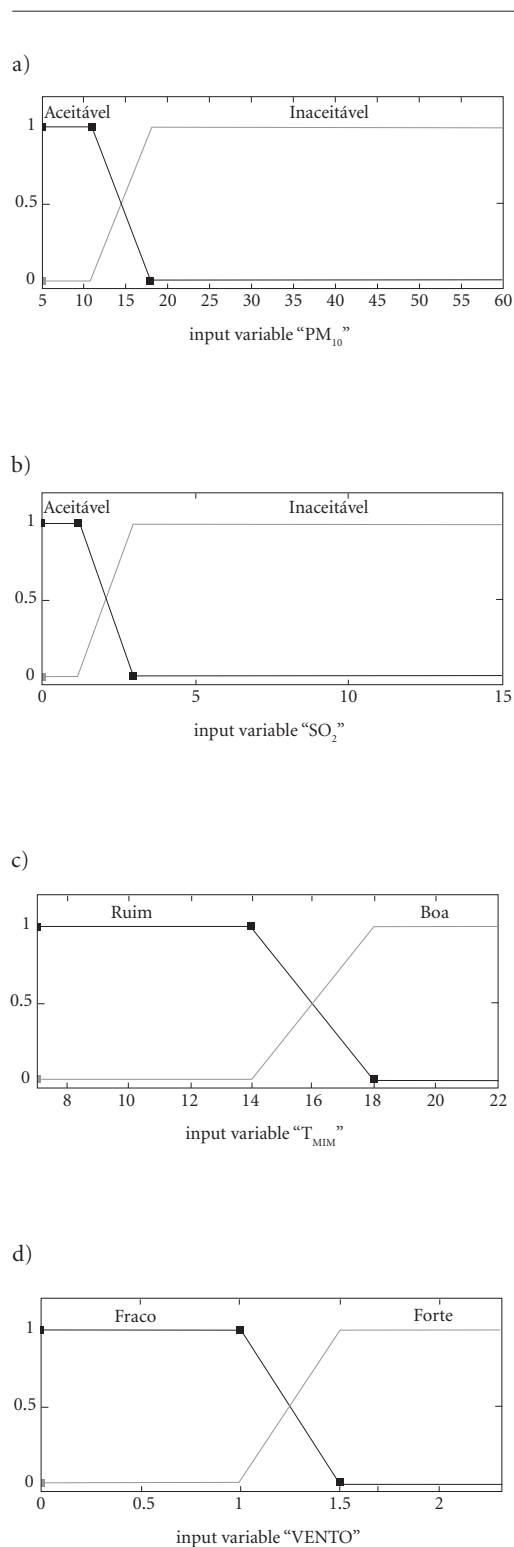


Figura 1. Variáveis de entrada: a) Material particulado (PM_{10}), b) Dióxido de enxofre (SO_2), c) Temperatura Mínima (T_{MIN}) e d) vento, para o modelo *fuzzy* para estimação do número de interações, São José dos Campos, 2009.

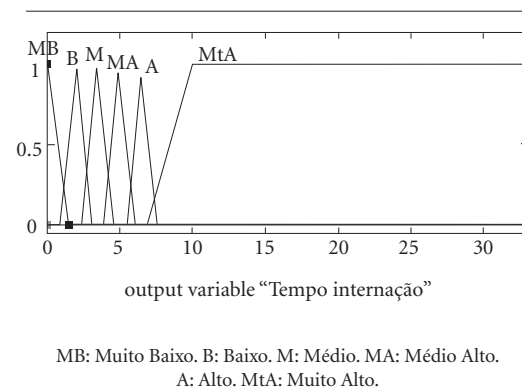


Figura 2. Variável de saída: tempo médio de internação por doenças cardiovasculares, São José dos Campos, 2009.

informações, o que impõe ao modelo maior flexibilidade e ajuste.

As partículas grandes de material particulado (diâmetro > 10 micrometros) são principalmente elementos derivados do solo, enquanto as partículas menores (diâmetro < 10 micrometros) são normalmente provenientes da queima de combustíveis fósseis e são perigosas por serem inaladas profundamente e atingir as vias respiratórias inferiores¹.

Concentrações típicas de PM_{10} , constatadas na América do Norte e na Europa Ocidental, se encontram na faixa de 20 a 50 $\mu g/m^3$, sendo que em áreas industrializadas esta faixa aumenta para 100 a 250 $\mu g/m^3$ ⁸.

Segundo dados da CETESB, a média aritmética das médias anuais de concentração do PM_{10} dos últimos 3 anos situa-se no intervalo de 20 $\mu g/m^3$ a 30 $\mu g/m^3$, enquanto para o SO_2 , a mesma encontra-se abaixo de 20 $\mu g/m^3$ ^{21,23}.

Neste período de estudo, somente em um dia foi considerada moderada a qualidade do ar para o PM_{10} e não houve ultrapassagem em nenhum dos dias no período considerado. Em relação ao SO_2 , não houve nenhuma ultrapassagem²¹. Em estudo anterior realizado em São José dos Campos²⁴ foram registradas concentrações médias de 40,2 $\mu g/m^3$ para o PM_{10} e 6,2 $\mu g/m^3$ para o SO_2 , o que denota uma diminuição de concentrações significativa, da ordem de 50%.

Como em estudos epidemiológicos é comum considerar-se os efeitos nocivos dos poluentes alguns dias após a inalação do mesmo, fez-se a correlação para diferentes defasagens (de zero a dois dias).

Tabela 1. Valores das médias, desvio padrão, mínimos e máximos das variáveis: material particulado (PM₁₀), dióxido de enxofre (SO₂), temperatura mínima (T_{MIN}) e vento, segundo tipo de saída tempo médio de internações, São José dos Campos-2009.

	Média	Desvio-Padrão	Mínimo	Máximo
PM ₁₀	21.62	8.76	7.00	57.00
SO ₂	2.90	1.85	0.63	15.00
T _{MIN}	15.89	2.83	7.58	20.96
Vento	0.94	0.36	0.20	2.30
Tempo médio de internações (real)	7.86	4.96	0.00	32.67
Tempo médio de internações (modelo)	14.42	6.89	1.42	20.82

Os dados obtidos pela curva ROC do modelo (Tabela 2) mostraram excelente sensibilidade, o PM₁₀ e o SO₂ obtiveram um bom desempenho para a defasagem zero, o que permite a conclusão de que os efeitos destes poluentes tornam-se mais evidentes no dia em que ocorre inalação dos mesmos. O melhor resultado foi para o poluente SO₂ na defasagem zero, com ótima acurácia e área sob a curva de 87,2%.

Devido à falta de estudos nesta área utilizando modelagem *fuzzy*, compromete-se a realização de comparações para uma conclusão mais precisa, não obstante, constam, na literatura nacional, artigos com aplicação *fuzzy* em: estimação do tempo de internação devido a pneumonia², devido a asma e pneumonia⁴, estabelecimento do risco de morte neonatal^{17,18} e previsão de reanimação neonatal²⁰.

Para a melhoria do modelo poderia ser adotado maior número de funções de pertinência, ou mesmo diferentes formatos como triangulares ou trapezoidais. As concentrações dos poluentes são consideradas homogêneas para a implementação desta abordagem, o que pode ser considerada uma limitação, visto que as concentrações em outras regiões da mesma cidade podem diferir daquelas registradas próximas à estação de monitoramento.

As melhores correlações foram obtidas para a defasagem zero (0,22) e para a defasagem de dois dias (0,23), o que mostra que o tempo de internação está relacionado à exposição aos poluentes do ar. Desta forma, quanto maiores as concentrações de PM₁₀ e SO₂ na atmosfera, é esperado maior tempo de internação de pacientes com doenças cardiovasculares.

Os dados reais do tempo médio de internação variaram de zero a 32,67 dias e os obtidos pelo modelo *fuzzy* variaram de 1,42 a 20,82 dias. A diferença entre os mínimos e os máximos se deve ao método de cálculo utilizado no sistema

Tabela 2. Valores para curva Roc e respectivos intervalos de confiança de 95%, para Lag 0, Lag 1 e Lag 2 dos poluentes PM₁₀ e SO₂, São José dos Campos, 2009.

	Lag 0	Lag 1	Lag 2
PM ₁₀	0,70 (0,61-0,77)	0,61 (0,54-0,67)	0,57 (0,50-0,67)
SO ₂	0,87 (0,83-0,93)	0,71 (0,64-0,77)	0,69 (0,63-0,77)

de inferência *fuzzy*, em que o resultado final é obtido pelo cálculo do centro de área, portanto seria matematicamente impossível a obtenção do centro de área próximo dos extremos do universo de discurso de cada variável.

Possíveis erros de diagnóstico, devidos à obtenção de dados de fontes cuja finalidade é apenas estatística e não epidemiológica, e embora inerente ao modelo, podem influenciar nos resultados finais. A não discriminação das internações, de acordo com gênero, pode ser outra limitação do estudo²⁵.

O modelo elaborado pode ser implementado de maneira eficiente em sistemas de saúde pública, podendo ser aplicado em quaisquer localidades onde existam dados disponíveis de poluentes e condições climáticas.

Este estudo apresenta uma ferramenta, de baixo custo financeiro, que possui a capacidade de prever, de forma satisfatória, a duração do tempo de internação por doenças cardiovasculares em consequência de variações climáticas e de concentrações de poluentes na atmosfera, assumindo um papel de significativa importância para que as equipes médicas e hospitalares possam estar preparadas para um atendimento mais rápido e eficaz.

Colaboradores

KMV Coutinho, PMSR Rizol, LFC Nascimento e APP Medeiros participaram igualmente de todas as etapas de elaboração do artigo.

Agradecimentos

À FAPESP pelo suporte financeiro.

Referências

1. Cançado JED, Braga ALL, Pereira LAA, Arbex MA, Saldiva PHN, Santos UP. Repercussões clínicas da exposição à poluição atmosférica. *J Bras Pneumol* 2006; 32(Supl. 1):S5-S11.
2. Nascimento LFC, Rizol PMSR, Peneluppi AP. Estimating the average length of hospitalization due to pneumonia: a fuzzy approach. *Braz J Med Biol Res* 2014; 47(11):977-981.
3. Braga ALF, Pereira LAA, Procópio M, André P, Saldiva PHN. Associação entre poluição atmosférica e doenças respiratórias e cardiovasculares na cidade de Itabira, Minas Gerais, Brasil. *Cad Saude Publica* 2007; 23(Supl. 4):S570-S578.
4. Chaves LE. *Modelos computacionais Fuzzy e Neuro-fuzzy para avaliarem os efeitos da poluição do ar* [tese]. Guaratinguetá: Universidade Estadual Paulista; 2013.
5. Cesar ACG, Nascimento LFC, Carvalho Jr JA. Association between exposure to particulate matter and hospital admissions for respiratory disease in children. *Rev Saude Publica* 2013; 47(6):1209-1212.
6. Gouveia N, Freitas C, Martins LC, Márcilio IO. Hospitalizações por causas respiratórias e cardiovasculares associadas à contaminação atmosférica no município de São Paulo, Brasil. *Cad Saude Publica* 2006; 22(12):2669-2677.
7. Nascimento LFC, Francisco JB. Particulate matter and hospital admission due to arterial hypertension in a medium-sized Brazilian city. *Cad Saude Publica* 2013; 29(8):1565-1571.
8. Mills NL, Donaldson K, Hadoke PW, Boon NA, MacNee W, Cassee FR, Sandström T, Blomberg A, Newby DE. Adverse cardiovascular effects of air pollution. *Nature Clinical Practice Cardiovascular Medicine* 2009; 6(1):36-44.
9. Martins LC, Pereira LA, Lin CA, Santos UP, Prioli G, Luiz OC. The effects of air pollution on cardiovascular diseases: lag structures. *Rev Saude Publica* 2006; 40(4):677-683.
10. Nascimento LFC. Air pollution and cardiovascular hospital admissions in a medium-sized city in São Paulo State, Brazil. *Braz J Med Biol Res* 2011; 44(7):720-724.
11. Brasil. Ministério da Saúde (MS). Portal da Saúde SUS. [acessado 2014 out 17]. Disponível em: <http://www2.datasus.gov.br/Datasus/index.php>
12. Amancio CT, Nascimento LFC. Environmental pollution and deaths due to stroke in a city with low levels of air pollution: ecological time series study. *São Paulo Med J* 2014; 132(6):353-358.
13. Yen J, Langari R. *Fuzzy logic: intelligence, control and information*. Upper Saddle River: Prentice-Hall; 1999.
14. Sellitto MA. Inteligência Artificial: uma aplicação em uma indústria de processo contínuo. *Gest Prod* 2002; 9(3):363-376.
15. Lima MMP, Murari CAF, Pereira MA. Uma nova metodologia para fluxos de carga fuzzy. *Sba: Controle & Automação Sociedade Brasileira de Automatica* 2003; 14(1):69-75.
16. Botura CA, Botura Junior G, Carvalho Jr JA, Mesquita L, Ferreira M A. Simulation of Active Control Using Fuzzy Logic Applied to a Pulse Combustor. *Journal of the Brazilian Society of Mechanical Sciences* 2002; 24(2):134-138.
17. Nascimento LFC, Rizol PMSR, Abiuzi LB. Establishing the risk of neonatal mortality using a fuzzy predictive model. *Cad Saude Publica* 2009; 25(9):2043-2052.
18. Nascimento LFC, Ortega NRS. Fuzzy linguistic model for evaluating the risk of neonatal death. *Rev Saude Publica* 2002; 36(6):686-692.
19. Zadeh LA. The concept of a linguistic variable and its application to approximate reasoning – I. *Information Sciences* 1975; 8(3):199-249.
20. Reis MAM, Ortega NRS, Silveira PSP. Fuzzy expert system in the prediction of neonatal resuscitation. *Braz J Med Biol Res* 2004; 37(5):755-764.
21. São Paulo. Companhia Ambiental do Estado de São Paulo (CETESB). Qualidade do ar no estado de São Paulo, 2012. [acessado 2014 set 30]. Disponível em: <http://www.cetesb.sp.gov.br>
22. São José dos Campos. Prefeitura Municipal. Portal Oficial. [acessado 2014 set 28]. Disponível em: <http://www.sjc.sp.gov.br>
23. São Paulo. Decreto Nº 59.113, de 23 de Abril de 2013. *Diário Oficial do Estado de São Paulo* 2013; 24 abr.
24. Nascimento LFC, Pereira LAA, Braga ALF, Módolo MCC, Carvalho Jr JA. Effects of air pollution on children's health in a city in Southeastern Brazil. *Rev Saude Publica* 2006; 40(1):77-82.
25. Clougherty JE. A growing role for gender analysis in air pollution epidemiology. *Cien Saude Colet* 2011; 16(4):2221-2238.

Artigo apresentado em 27/11/2014

Aprovado em 27/04/2015

Versão final apresentada em 29/04/2015