DOI: 10.1590/1413-81232015208.19472014

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*

[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.

Introduction

Several epidemiological studies have shown air pollution as one of the factors that has the greatest impact on human health, mainly associated with diseases of the respiratory[1-6] and circulatory system[6-10]. In cities that are important industrial centers, such as São José dos Campos, population faces serious environmental problems due to cars and trucks heavy traffic, such as high accumulation of air pollutants, high population density and major industrial development.

According to the Department of Information and Informatics of the Unified Health System (Datasus) in 2012 1.1 million hospital admissions were recorded due to cardiovascular disease in Brazil and 600,000 in São Paulo, generating an expenditure of R$2.4 billion (1US$ ≈ R$2.50) to the public funds[11].

Particulate matter (PM_{10}), sulfur dioxide (SO_2), nitrogen oxides (NO and NO_2) and ozone (O_3) are the pollutants studied and the most commonly associated with deleterious effects on human health[1]. PM_{10} is a mixture of solid and liquid particles with a diameter less than 10μm, suspended in the air and produced by fuels and biomass burning; SO_2 is generated from the combustion of fossil elements as diesel oil and O_3 is formed by reaction between the ultraviolet radiation, nitrogen oxides and hydrocarbons emitted by vehicles[12].

Studies show that individuals exposed to high pollution levels have increased blood pressure, heart rate variability, which provides greater risk of arrhythmia and sudden death, and heart attack and stroke, and a high incidence of bronchial inflammation, accompanied by severe decline pulmonary function[2,4,7,10].

It is believed that the particulate matter (PM_{10}) and sulfur dioxide (SO_2) are closely related with hospitalizations and deaths due to heart failure, risks that are more powerful in the day that occurs exposure to pollutants[7].

Statistical techniques such as logistic regression and Poisson regression (Generalized Linear Models or Models-MLG Generalized Additive-MAG) are currently used in many studies in this area[5]. These models are able to estimate the chance or the risk of hospitalization or death due to respiratory and cardiovascular diseases.

Nevertheless, fuzzy models emerge as a new tool of choice due to their ability to deal with vagueness and uncertainty of information[13], where values can be classified as partially true, has ease of understanding and low computational cost. This approach has been used in various fields such as in industrial process control[14-16] and medicine[2,4,17,18].

The concept of fuzzy logic was proposed by Lotfi Zadeh and its theory is based on concepts of classical logic, however with the definition of pertinence degrees[13,19].

In classical theory, the sets are called crisp, where an element of a universe of discourse may or may not belong to a certain set. In the theory of fuzzy sets, as there is a degree of pertinence of each element to a certain set, there is no border from which you can define whether an element belongs to the set in question, and so, it became an important mathematical approach to be applied in the health field, probably due to its ability to cope with linguistic terms.

In this area, fuzzy logic has been used in modeling in the process of X-ray diagnosis, fuzzy standards of recognition and for X-ray photos analysis in epidemiology and public health[2,4,17,18,20].

The aim of this study is to build a computer model using the fuzzy logic properties to estimate the mean hospitalization time due to cardiovascular diseases according to air pollutants concentrations in São José dos Campos, Brazil.

Methods

This is a computational model using fuzzy logic to estimate the hospitalization mean time due to cardiovascular diseases, according to concentrations of PM_{10} (particulate matter), SO2 (sulfur dioxide), minimum temperatures values and wind speed obtained the Environmental Company of São Paulo State (Cetesb), which has a metering station in São José dos Campos - SP[21].

The model was developed based on real data of admissions for cardiovascular disease in individuals of all ages living in the city of São José dos Campos, in the period from 01/01/2009 to 31/12/2009. These data were obtained from Datasus, referring to the diagnostics of chapter IX.

São José dos Campos is a medium-sized city in the state of São Paulo which has an important industrial park. Located in Alto Vale do Paraíba, at 660m above sea level, it has damp and tropical climate, being located between Sao Paulo and Rio de Janeiro - the two largest cities in Brazil; It is crossed by the Dutra Highway, the Brazilian most important road that has heavy traffic due to buses and trucks throughout the year, about 130 000 vehicles per day. Its population is estimated to be around 700 thousand inhabitants[22].

This model is based on the fuzzification of the variables particulate (PM_{10}), sulfur dioxide (SO_2), with minimum temperature and wind speed.

The model was developed with the help of an expert who developed two membership functions, Acceptable and Unacceptable to the input variables: particulate matter (PM_{10}), sulfur dioxide (SO_2), minimum temperature (T_{MIN}) and Weak and Strong to the input variables wind speed. The output was the mean time of hospitalization, with six membership functions: very low, low, medium, medium high, and very high. The universes of discourse in which the contents of each variable were considered low or high were stipulated with basis on how harmful to health the element in question is.

On combining all possible inputs, each one with two pertinence functions, the construction of 16 rules was possible (2x2x2x2).

The fuzzy linguistic model procedure consists in calculating the degree of pertinence of the values PM_{10}, SO_2, T_{MIN} and wind speed. Then it is performed the fuzzy inference process proposed by Mamdani; the number of hospitalizations is estimated by defuzzification technique based on the center area method, as described in a previous article[17].

Through the routine using the MATLAB® program, it is possible to generate a numerical output to the fuzzy model, therefore the mean hospitalization time was obtained from the input values of each variable. From these responses the hospitalization mean time obtained is compared to the real data hospitalizations, using a lag of zero up to two days.

For the evaluation of the results Pearson's correlation was made and the accuracy was determined using the ROC curve.

Due to the fact that the data available were on the net and not identified the submission to the Ethics Committee on Research was dismissed.

Results

During this period 1119 hospitalizations were registered in the city of São José dos Campos due to cardiovascular disease diagnosis.

The input variables: particulate matter (PM_{10}), sulfur dioxide (SO_2), minimum temperature (TMIN) and wind, are shown in Figure 1 (a-d). The output variable: hospitalization mean time due to cardiovascular diseases, with their respective membership functions, is shown in Figure 2.

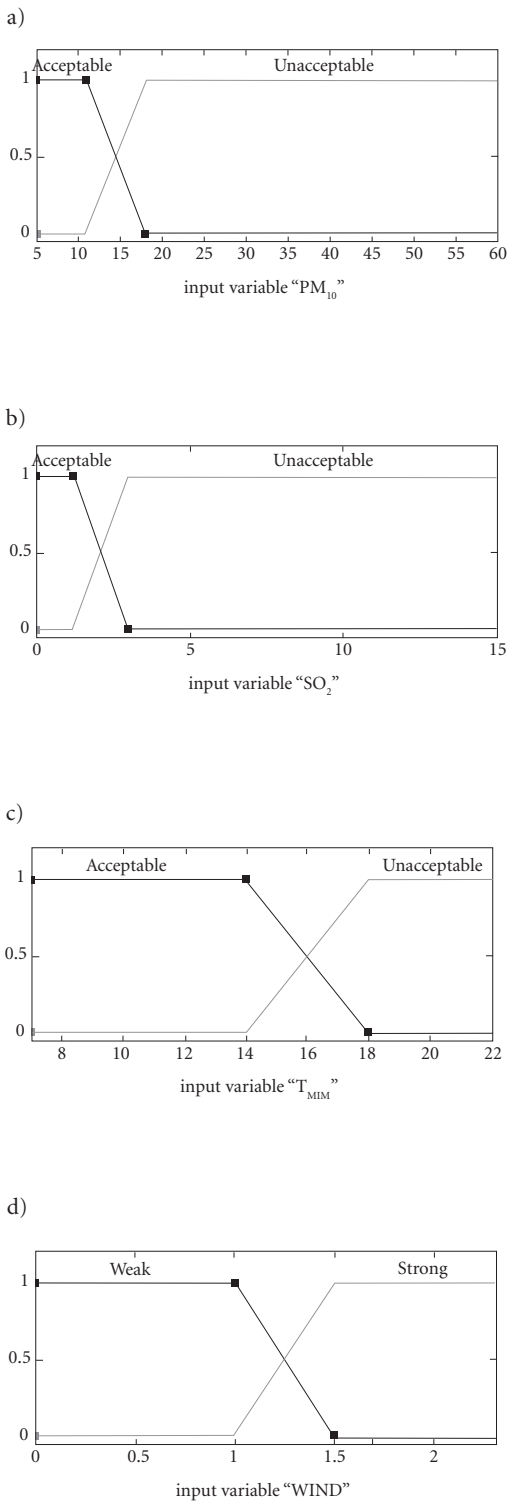
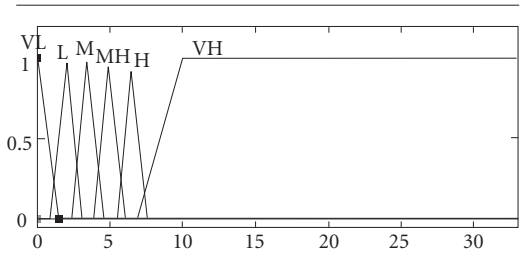


Figure 1. Input variables: a) Particulate matter (PM_{10}), b) Sulfur dioxide (SO_2) c) Minimum Temperature (TMIN) and d) Wind for the fuzzy model to estimate the number of hospitalizations, São José dos Campos, 2009.



output variable "Lenght of hospitalization"

VL: Very Low. L: Low. M: Medium. MH: Medium High. H: High. VH: Very High.

Figure 2. Output variable: mean hospital stay for heart disease, São José dos Campos, 2009.

Mean values, standard deviations, minimum and maximum values of variables particulate matter (PM_{10}), sulfur dioxide (SO_2), minimum temperature, wind speed and mean time of hospitalization are shown in Table 1.

The results provided by the ROC curve are shown in Table 2. The best result was for the lag zero, PM_{10} and SO_2 showed best performance by the ROC curve, with statistically significant values.

The Pearson correlation between the results provided by the fuzzy model and the real data was 0.22 ($p < 0.01$) to lag 0, 0.15 to lag 1 and 0.23 to lag 2, although not being high values, they are significant.

Discussion

As far as is known, this is the first study conducted in Brazil involving the effect of exposure to air pollutants on the hospitalization mean time due to cardiovascular disease using fuzzy logic tool.

This study deals with the construction of a computer model with fuzzy approach to estimate the hospitalization mean time due to cardiovascular disease related to air pollutants in an industrial medium-sized city, showing good accuracy to predict the hospitalization mean time when exposure occurred on the same day, even on day up until two days later.

Studies on air pollution impacts on human health usually make use of tools such as logistic regression and Poisson, where the patients length of hospitalization is generally not considered; nevertheless, this constitutes a variable of extreme importance due to the fact that it represents the actual cost to society.

In relation these models, the fuzzy approach has the advantage of dealing with the uncertainty of information, which imposes larger flexibility and adjustment to the model.

Large particles of particulate matter (diameter> 10 microns) are mainly soil derived elements, while smaller particles (diameter <10 microns) are usually derive from fossil fuels burning and are dangerous when inhaled because it can achieve more distal portinons of lungs[1].

Typical concentrations of PM10, found in North America and Western Europe, are in the 20-50 μg/m^3 range, and in industrialized areas this range increases to 100 to 250 μg/m^3 [8].

According to Cetesb data basis, the arithmetic average of the annual average PM10 concentration of the last three years lies in the range of 20 μg/m^3 to 30 μg/m^3, while for SO2, it is below 20 μg/m^3 [21,23].

In this study period, only one day was considered to have moderate air quality for PM10 and no change on any of the days in the period considered. Considering SO2, there was no change[21]. In a previous study conducted in São José dos Campos[24] average concentrations were recorded as 40.2 μg/m^3 for PM10 and 6.2 mg/m^3 for SO2, which shows a decrease of significant concentrations of around 50 %.

As in epidemiological studies it is common to consider the pollutants harmful effects a few days after its inhalation, a correlation was made for different time lags (from zero to two days).

The data model obtained by the ROC curve (Table 2) showed excellent sensitivity, PM_{10} and SO_2 obtained good performance for zero lag, what enables the conclusion that the effects of these pollutants become more evident on the day their inhalation happens. The bests result was for SO_2 pollutant at zero lag, with great accuracy and area under of an 87.2% curve.

Due to lack of studies in this area using fuzzy modeling, the making of comparisons is compromised as to a more precise conclusion, nevertheless, there are in Portuguese articles with fuzzy applications on: estimate of hospitalization duration due to pneumonia[2], due to asthma and pneumonia[4], establishment of neonatal death[17,18] and prediction of neonatal resuscitation[20].

To improve the model a greater number of membership functions could be adopted, or even different formats such as triangular or

Ciência & Saúde Coletiva, 20(8):2585-2590, 2015

Table 1. Mean values, standard deviation, minimum and maximum of the variables: particulate matter (PM_{10}), sulfur dioxide (SO_2), minimum temperature (T_{MIN}) and wind, according to type of output mean time of hospitalizations, São José dos Campos-2009.

	Average	Standard deviation	Minimum	Maximum
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
Wind	0.94	0.36	0.20	2.30
Average length of hospitalization (real)	7.86	4.96	0.00	32.67
Average length of hospitalization (model)	14.42	6.89	1.42	20.82

trapezoidal ones. The pollutant concentrations are considered homogeneous to implement this approach, what can be considered a limitation, since concentrations in other regions of the same city may differ from those recorded close to the monitoring station.

The best correlations were obtained for zero lag (0.22) and for the two-day lag (0.23), which shows that the hospitalization period is related to exposure to air pollutants. Thus, the higher the PM_{10} and SO_2 concentrations in the atmosphere, the longer hospitalization period of patients with cardiovascular disease is expected.

The real data of hospitalization average period ranged from zero to 32.67 days and those obtained by the fuzzy model ranged from 1.42 to 20.82 days. The difference between the minimum and maximum values is due to the calculation method of the fuzzy inference system, where the final result is obtained by the area center calculation, therefore it would be mathematically impossible to obtain the area center near the discourse universe extremes of each variable.

Possible misdiagnosis, due to obtaining sources of data whose purpose is not just statistical and epidemiological, and although inherent to the model, they may influence the final results. Non-discrimination of admissions, according to gender, can be another limitation of the study[25].

The model developed can be implemented efficiently in public health systems and can be applied in any locations where available data of pollutants and climatic conditions exist.

This study presents a tool of low financial cost, which can satisfactorily predict, the duration of hospitalization due to cardiovascular disease as a result of climatic variations and concentrations of pollutants in the atmosphere, assuming a role of significant importance so that medical staffs and hospitals can be prepared for a faster and effective service.

Table 2. Values for ROC curve and 95% confidence intervals to 0 lag, lag 1 and lag 2 of PM_{10} and SO_2 pollutants, 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)

Collaborations

KMV Coutinho, PMSR Rizol, LFC Nascimento and APP Medeiros participated equally of all stages of the article preparation.

Acknowledgement

To São Paulo Research Foundation (FAPESP) for financial support.

References

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, Marcílio 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.

Article submitted 27/11/2014
Approved 27/04/2015
Final version submitted 29/04/2015