ARTICLE IN PRESS

Atmospheric Pollution Research xxx (xxxx) xxx–xxx

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Atmospheric Pollution Research

journal homepage: www.elsevier.com/locate/apr

Negative Binomial regression model for analysis of the relationship between hospitalization and air pollution

Leda G. Ardiles[a], Yara S. Tadano[b], Silvano Costa[c], Viviana Urbina[a], Maurício N. Capucim[a], Iara da Silva[a], Alfésio Braga[d], Jorge A. Martins[a], Leila D. Martins[a,*]

[a] Federal University of Technology - Parana, Av. Dos Pioneiros, 3131, Londrina, 86047-125, Brazil
[b] Federal University of Technology - Paraná, Mathematics Department, Ponta Grossa, 84016-210, Brazil
[c] Department of Statistics, State University of Londrina - UEL, PR 445 - Km 380, Caixa Postal 6001, CEP 86051-990, Londrina, Brazil
[d] Group of Environmental Exposure and Risk Assessment Studies, Post-Graduate Program in Public Health, Universidade Católica de Santos, Santos, SP, Brazil

A R T I C L E I N F O

Keywords:
Air pollution
Health effects
Brazilian cities
Best regression
Emission sources

A B S T R A C T

Air pollution is currently one of the major environmental risk to human health in the world. Several studies link health to air pollution through the Poisson regression model. However, this study showed that the best method is the regression model with Negative Binomial distribution. The analysis was made based on daily data for the period between 2010 and 2014 regarding hospitalizations due to circulatory and respiratory diseases, stratified by age due to the concentration of air pollutants: CO, PM_{10}, NO_2, SO_2 and O_3 and meteorological variables. The results with Negative Binomial distribution model indicate, for interquartile range (IQR), that CO is a risk factor (1.00259–1.19688, 95% CI) for circulatory diseases in the elderly (aged over 64) and for respiratory diseases in the population aged 10–19 years old, while NO_2 is a risk factor for the population in the range of 20–64 years (1.00313–1.25827, 95% CI) in the city of Araucaria. In Curitiba, air pollutants presenting a risk factor for respiratory diseases were PM_{10} (RR = 1.05322) in the range of 0–9 years old, CO and O_3 (RR = 1.05047) in the range of 20–64 years old, and CO for the elderly. It could be verified that exposure to air pollutants is harmful to health even when their concentrations are below the standards set out in legislation. Additionally, the city of Araucaria presented the highest number of hospital admissions, which may be associated with the large number of industries and lower human development index in the city.

1. Introduction

Air pollution has been investigated globally, mainly because of its risks to human health, since it causes severe respiratory and cardiovascular problems, among others, and could even lead to death. Urban areas are the most affected due to the high amount of pollution sources and number of people living there. It is estimated that throughout the world, this type of pollution is responsible for the death of 7 million people each year (WHO, 2014). Air pollution is currently a major environmental risk to health worldwide (WHO, 2014), therefore, scientific studies on the subject are relevant.

The adverse health effects are attributed both to acute and to chronic exposures, and different pollutants in the atmosphere can cause a great variety of adverse health effects (Künzli et al., 2010; Hime et al., 2015; Ostro et al., 2015). The most frequent adverse effects occur in the respiratory and circulatory system in the elderly (Saldiva et al., 1995; Arbex et al., 2012; Franklin et al., 2015; Weuve et al., 2016), patients with pre-existing chronic diseases, children, and during pregnancy, who make up the most vulnerable group to air pollution (Schwartz, 2004; Magas et al., 2007; Kim et al., 2015; Deng et al., 2016, 2017).

Environmental pollution causes increased morbidity and mortality, with risk to the entire population, because anyone can be exposed to these contaminants. Experimental and epidemiological investigations are the main tools for understanding the health effects caused by air pollution. A strategy for estimating these health effects associated with exposure to the concentration of pollutants and meteorological variables is the use of Generalized Linear Models (GLM), introduced by Nelder and Wedderburn (1972), as a proposed expansion to the classic linear models.

Several studies have used the Poisson distribution to assess the impact of air pollution on human health, (i.e. Fischer et al., 2011; Vanos et al., 2014; Çapraz et al., 2016; Samoli et al., 2016; among others), and other have been apply the Negative Binomial distribution model (Morris et al., 1995; Magas et al., 2007; Rozita et al., 2010; Leiva G

Peer review under responsibility of Turkish National Committee for Air Pollution Research and Control.
* Corresponding author.
E-mail address: leilamartins@utfpr.edu.br (L.D. Martins).

http://dx.doi.org/10.1016/j.apr.2017.10.010
Received 8 June 2017; Received in revised form 26 October 2017; Accepted 27 October 2017
1309-1042/

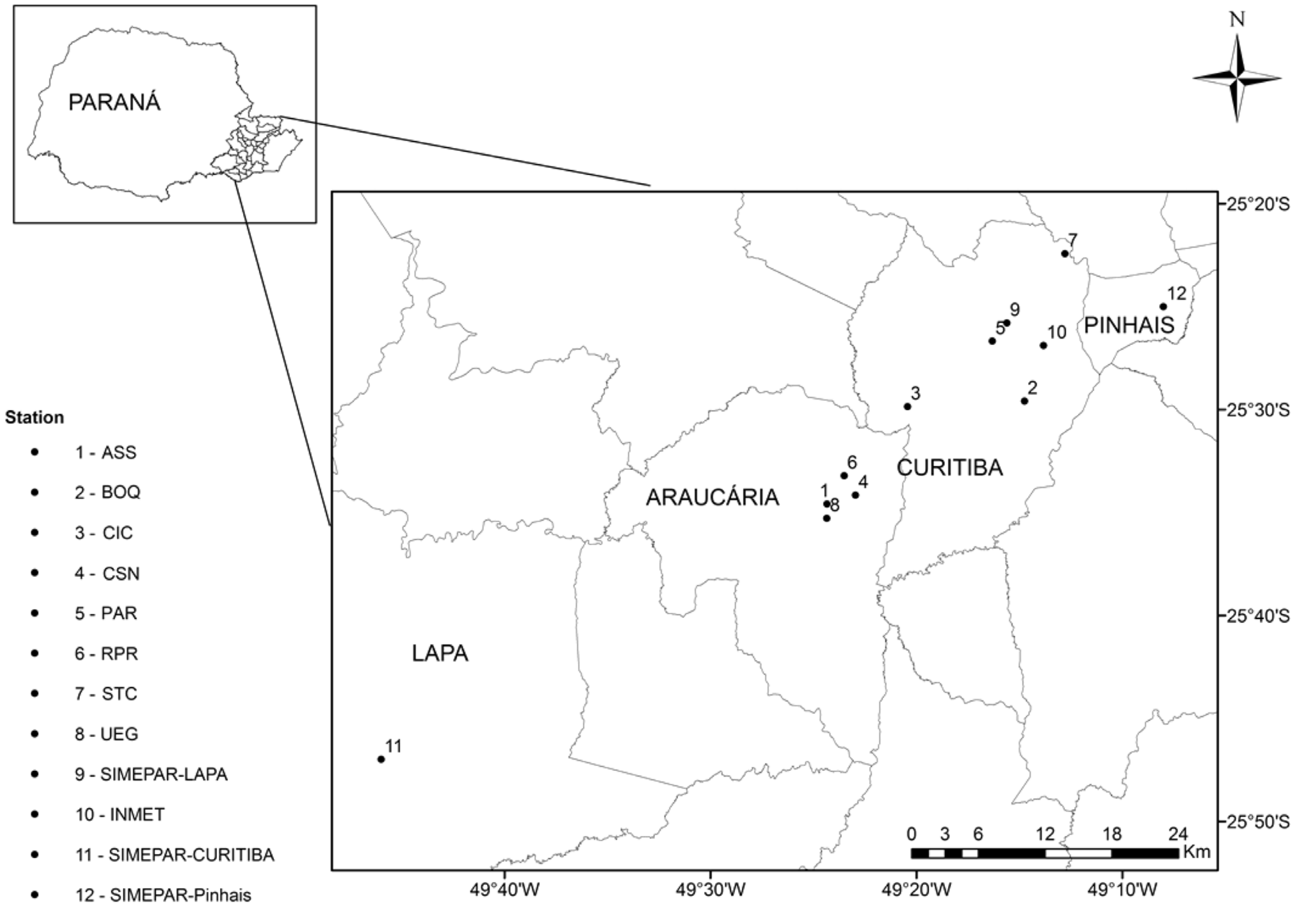

Fig. 1. Location of the study areas, with air quality (1–8) and meteorological (9–12) stations.

et al., 2013; Yaser et al., 2014). However, are scarce the works that previously analyzed the fit of distribution to dataset before to analyzed the associations. Thus, the objective of this study was to show the best distribution model for this type of dataset before apply it to investigate the impact of CO, PM_{10}, NO_2, SO_2 and O_3 concentrations and meteorological variables (temperature and relative humidity) recorded in two cities - with significant differences on emission amount and source types - on the number of daily admissions due to respiratory and circulatory diseases.

2. Methodology

2.1. Study area

The study was conducted in the municipalities of Araucaria with 119,123 inhabitants and Curitiba, with 1.75 million inhabitants (IBGE, 2010), in the period from 01/01/2010 to 31/12/2014. Fig. 1 shows the location of the study sites - Curitiba and Araucaria, as well as the air quality and weather monitoring stations from which data were used for the analysis.

This local configuration of the air quality monitoring stations is due to the different characteristics of pollutant sources in Curitiba and Araucaria. The predominant source in Curitiba is vehicular, while in Araucaria, with a much smaller urban area, is industrial. Araucaria has higher total emission of pollutants when compared with Curitiba, being around 1.6, 26 and 1.9-fold higher for particles, SOx and NOx, respectively, with most of the emissions (95%) being from industrial sources (IAP, 2013).

In the study region, the highest temperatures and rainfall occur during the summer months, that is, from December to February, and the lowest temperatures are recorded between May and August. Rainfall occurs throughout the year, but less intensely during the winter, with this season being considered cold and not as humid, making the atmosphere unfavorable for the dispersion of pollutants in this period. The climatological average monthly rainfall ranged from 73 to 190 mm $month^{-1}$; the monthly relative humidity ranged from 76% to 82%, and the temperature between 8 °C and 27 °C. The wind direction during the studied period was predominantly from the East. The wind is generally weak, with an average speed between 2 and 4 m s^{-1}.

2.2. Dataset used

In order to analyze which distribution model is the best fit to correlate air pollutant concentrations (CO, PM_{10}, NO_2, SO_2 and O_3), meteorological variables (pressure (P_m), relative humidity (RH_m) and temperature difference ($T-T_m$)) with daily number of hospital admissions due to circulatory and respiratory diseases, two dataset, one from Curitiba and one from Araucaria, were used. The datasets are from 1 January 2010 to 31 December 2014.

The pollutant hourly concentrations were provided by Lactec Institute for the eight automatic stations, located in Curitiba and Araucaria (Fig. 1), which have measurements available. Table 1 identify the air quality stations and the pollutants measured and analyzed in each stations as well as the place type of each station. The average daily concentrations of each pollutant were used for the regression analysis to stations located in Araucaria and Curitiba. Both cities presented 100% of stations and variables valid, since they have more than 75% of hourly data valid in the day.

The meteorological variables were obtained from the National Institute of Meteorology (INMET) and Parana Meteorological System (SIMEPAR) for the stations located in the study area. For regression analysis, were used the daily average between INMET and SIMEPAR