Korean Journal of Clinical Medicine

Review Article Open Access

Modeling of Hospital Admissions for Respiratory Diseases as a Function of Probability Distribution Functions

Souza A[1*], Abreu MC[2], de Oliveira-Júnior JF[3], Cavazzana G[4], dos Santos CM[5], Pobocikova I[6], Fernandes WA[1], Garcia Oliveira AP[7] and Barros da Silva E[8]

[1]Department of Medicine,Federal University of Mato Grosso do Sul, C.P. 549, 79070- 900. Campo Grande, MS – Brazil

[2]Department of Medicine, Universidade Federal Rural do Rio de Janeiro, Seropédica, Rio de Janeiro, Brasil

[3]Department of Medicine, Universidade Federal de Alagoas, Instituto de CiênciasAtmosféricas (ICAT), Maceió, Brazil

[4]Department of Medicine, Universidade Católica Dom Bosco and Faculdade de Mato Grosso do Sul, Brazil

[5]Department of Medicine , Federal University of Para, Altamira, PA, Brazil

[6]Department of Applied Mathematics, Faculty of Mechanical Engineering, University of Zilina, Slovakia

[7]Department of Medicine, Universidade Anhanguera-Uniderp, Campo Grande, MS – Brazil

[8]Department of Medicine , Secretaria Municipal de Capela (SMC),Capela , 57780-000 , Alagoas, Brasil

***Corresponding author:** Amaury de Souza, Department of Medicine, Federal University of Mato Grosso do Sul, C.P. 549, 79070- 900. Campo Grande, MS – Brazil, E-mail: amaury.de@uol.com.br

Citation: Souza A (2020) Modeling of Hospital Admissions for Respiratory Diseases as a Function of Probability Distribution Functions . Korean Journal of Clinical Medicine. V1(2): 1-7.

Received Date: Mar 28, 2020 **Accepted Date**: Apr 20, 2020 **Published Date**: Apr 27, 2020

1. Abstract

1.1. Objective: To analyze the adjustments of the weibull, gamma, normal and logistic probability density distributions of the historical series of hospitalizations for respiratory diseases (childhood and adult pneumonia) from 2011 to 2015, in Campo Grande, MS.

1.2. Methods: The shape and scale parameters of the distributions were determined to verify the quality of the data fit.

1.3. Results: The R2, MAE, RSME, MAPE tests were used to verify the best estimate of hospitalization data.

1.4. Conclusion: The best fit was the Gamma distribution these distributions can be used as an alternative distribution that adequately describes the data on hospital admissions for respiratory diseases in Campo Grande.

2. Keywords: Hospital admission; Pneumonia: Modeling; Probability; Child, and Adults.

3. Introduction

Several studies of adjustment of probability density distribution or probability estimates using theoretical models of probability in relation to a historical series of data have been developed, highlighting the benefits in the planning of activities that minimize the risks, among which can be cited: precipitation [1-5], air temperature [6,7], concentration of pollutant gases [8,9], for the historical series of hospital admissions for respiratory diseases there are no published works with this methodology.

The use of probability density functions is directly linked to the nature of the data to which they relate. Some have good estimation capacity for small numbers of data, others require a large number of observations. Due to the number of parameters of your equation, some can take different forms, being framed in a greater number of situations, that is, they are more flexible. Since respecting the aspect of data representativeness, the estimates of its parameters for a given region can be established as general purpose, without prejudice to the precision in the estimation of probability [1].

Climate change has become one of the mostserious environmental concerns for urban areas in recent decades. Several epidemiological studies in recent years have reported associations between high levels of climatic changes and increased rates of death and hospitalization for respiratory and cardiovascular diseases [10,11]. Some epidemiological studies show that air pollution affects human health, even concentrations of air pollutants are below the air quality standards [12].

Respiratory diseases and related mortality have been increasingly associated with exposure to climate change. Sensitive and vulnerable groups, such as pregnant women, children, the elderly, and those who already suffer from respiratory illnesses and other serious diseases, or from low-income groups, are especially affected by climatic variation. Studies have shown that the number of respiratory diseases in children and the elderly increases due to the higher concentrations of air pollution[13-19]. According to these studies, children are more susceptible because they need twice the amount of air inhaled by adults, and the elderly are more affected because of their weakened immune and respiratory systems and have been exposed to a large amount of air pollution in all your life.

There are no published works (according to the best knowledge of the authors) on historical series of hospitalizations for respiratory diseases based on the methodology used in this research, developed by [20], which analyzed the adjustments of the Burr (Bu), Inv Gausian 3P (IG3P), Lognormal (LN), Pert (Pe), Rayleigh 2P (Ra 2P) and Weibull 3P (W3P) distributions in the series of hospitalizations for respiratory diseases (total hospitalizations)., the parameters of the shape and scale of the distributions were determined and, to check the quality of the adjustment of the observation data, the quality adjustment tests (GOF) were applied: Kolmogorov-Smirnov, Anderson test - Darling Test, chi-square tests.

In the present study, which is the first hospital admission modeling study in Campo Grande, MS, Brazil, we performed a daily time series study of hospital admissions for respiratory conditions in Campo Grande, using probability distribution functions, Weibull, Gamma, Normal and logistics. The analysis was performed between people (children and adults) who contracted pneumonia.

4. Methods

For the modeling of the data of hospital admissions in Campo Grande weused the functions of Weibull, Gamma, Normal and Logistica. Performance indicators are calculated by comparing observed values to predicted values. The observed values are the classified values of the observation data, while the predicted values are the values obtained from the adjusted distribution**.**

4.1. Data Collection

The city of Campo Grande - MS, (20° 27'16 "S, 54° 47'16" W, 650 m), islocated on the plateau called Maracaju-Campo Grande, 150 miles from the start of the largest flood plain in the world, the Pantanal (139 111 km2), and an estimated population of 724,000 inhabitants. Souza & Granja [21], using the Koppen's method, the climate in the region of Campo Grande, is the type with moderate temperatures ranging from 17.8°C minimum, 29.8°C maximum and average of 22.7°C, with hot summer and well distributed rainfall, average relative humidity is 72.8%, found prevailing winds in East Campo Grande - MS, occurring in the North in months from January to December, with annual values resulted in 24% of East, 19.8% of North and 12,2% of Northeast, and the lulls represented 12% with an average speed of 3.1 m/s, and average monthly rainfall in 122, 4mm and annual average 1469 mm.

For the correlation of weather data with the aggravation of respiratory illnesses, hospitalization data were collected from the health agencies of SUS (Unified Health System) and Department of Informatics (DATASUL).

The available data came from the Hospital Information System of SUS (SIH / SUS), managed by the Ministry of Health, through the Department in Health Care, in conjunction with the State Departments of Health and the Municipal Health and processed by Datasus at the ExecutiveDepartment of the Ministry of Health.

All hospitalizations occurred in the period between January 1st, 2011 and December 31st, 2015, the diseases investigated were coded according to the International Classification of Diseases (CID) 10th Revision, Pneumology (J17). The subjects of this study were children under years of age and adult.

Table 1: Geographical coordinates of the measurement site

City	Latitude	Longitude	Altitude (m)	Area (KM^2)	Measuring period
Campo Grande	20º26'34"S	54º38'47"W	532	8118,4	Jan to Dec 2013/2015

4.2. Probability Distributions

In this study the effectiveness of twelve one-component probability distributions are evaluated. We have used the one-component parametric pdfs because our data presents a unimodal distribution. These twelve models have been selected among other one-component models due to their successful applications according to the literature. The used pdfs as well as their cumulative distribution functions (cdfs) and the number of their parameters (N) are presented in Table 2.

Table 2: List of used pdfs, their cdfs and number of parameters.

Name	pdf	cdf	N	Nomenclature
W2	$f(v)=(k/c)\,(v/c)^{(k-1)}\exp(-(v/c)^k)$	$F(v)=1-\exp(-(v/c)^k)$	2	c is the scale parameter and k is the shape parameter
N	$f(v)=1/(\sigma\sqrt{2\pi})\exp(-(v-\mu)^2/(2\sigma^2))$	$F(v)=1/2\,(1+\mathrm{erf}((v-\mu)/(\sigma\sqrt{2})))$	2	μ = mean (location parameter) σ= variance (squared scale)
G	$f(v)=v^{(\alpha-1)}/(\Gamma(\alpha)\,\beta^{\alpha})\exp(-v/\beta)$	$F(v)=(\gamma(\alpha,v/\beta))/\Gamma(\alpha)$	2	α, β, Γ and γ are the shape parameter, scale parameter, the gamma function, And the incomplete Gamma function respectively
L	$f(v)=(\exp(-(v-\mu)/\sigma))/(\sigma\,[1+\exp(-(v-\mu)/\sigma)]^2)$	$F(v)=1/(1+\exp((v-\mu)/\sigma))$	2	μ is the location parameter and s is the scale parameter.

4.3. Estimation of Distributions Parameters

Several methods can be used to estimate the considered distributions parameters [22]. However, the selection of effective distributions is more important compared to the selection of parameter estimation methods [23]. In this work, the Maximum Likelihood method (ML) is applied. This method has shown good results in several studies. It gives the values of the parameters which maximize the probability of obtaining the observed data.

The likelihood function (L) for a random sample of wind speed v_1, v_2, …, v and theoretical probability density function (f) with j parameters $\alpha_1, \dots, \alpha_j$ is represented by equation:

$$L = \prod_{i}^{n} f(v_i, \alpha_1, \dots, \alpha_j) \quad (1)$$

For each parameter α_i, ML consists in estimating its value which maximizes the Likelihood function (L) by solving the following equation.

$$\frac{d \log L}{d\alpha_i} = 0 \quad (2)$$

4.4. Accuracy Tests

The accuracy tests (or goodness of-fit tests) are essential to compare the observed climate distributions with the predicted/ modelled distributions. The observed dataset is the values from the monitoring systems where as the modelled datasets are obtained from the fitted distributions. In this study, two categories de goodness of-fit tests are used. R^2 and RMSE associated with pdf (which are calculated using the relative frequencies of the histogram and the predicted pdfs obtained by theoretical model) and MAE and MAPE associated with cdf (which are calculated using the empirical cumulative frequencies of observations and the predicted cdf of the studied models). These accuracy tests are based on histogram approach, in which the measurements are arranged in a relative frequency histogram with N class intervals. The advantage of this approach is that it is less affected by individual measurements [24]. Their expressions are given below.

4.5. The Coefficient of Determination (R^2)

The coefficient of determination measures how much the variance of the measured data is explained by the theoretical model. In this work. R^2 is calculated using at the class intervals the relative frequencies of the histogram and the predicted pdf**s** obtained by theoretical model [24]. R^2 is expressed as follow:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(p_i - \hat{p}_i)^2}{\sum_{i=1}^{N}(p_i - \bar{p})^2} \quad (3)$$

where $\hat{p}_i$ is the predicted pdf at the at ith interval, p_i is the relative frequency at the at i^{th} class and $\bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i$ [25].

4.6. Root Mean Square Error (RMSE)

Since, it combines the bias and the dispersion, the root means square error is an important indicator for comparing the predicted with the observed values. The RMSE associated with probabilities in class intervals is given as:

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}(p_i - \hat{p}_i)^2\right]^{\frac{1}{2}} \quad (4)$$

Mean Absolute Error (MAE)

The mean absolute error is defined as the mean of the absolute errors derived from the observed and predicted values. The mathematical equation of MAE associated with cdfin class intervals is defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|F_i - \hat{F}_i\right| \quad (5)$$

where $\hat{F}_i$ is the theoretical cdf of the i^{th} measured wind speed and F_is the empirical cdf of the measured wind speed at i^{th} time stage.

Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error indicates the mean absolute percentage difference between the predicted and observed data. Basing on the histogram approach, the mean absolute percentage error associated to the cdfs is calculated as [26]:

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|F_i - \hat{F}_i|}{F_i}\times 100\% \quad (6)$$

4.7. Ethical Considerations

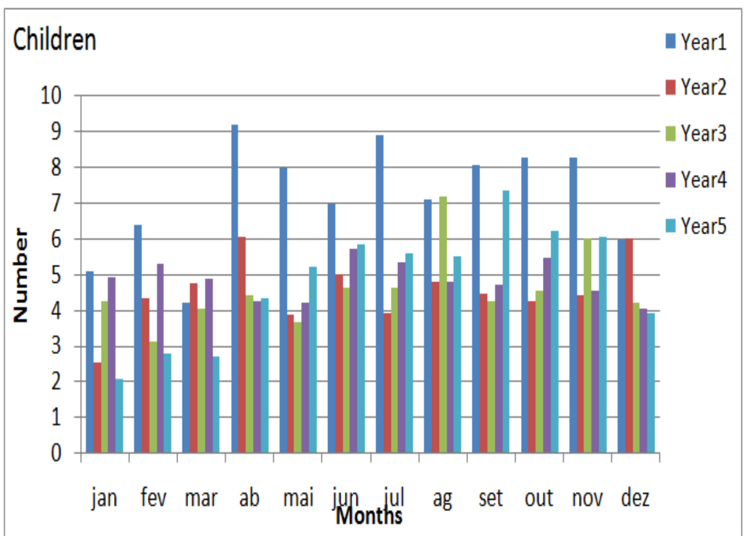
The present study is based on secondary, publicly available data, which do not constrain groups of populations and / or individuals in the presentation of the results found, ensuring the confidentiality of the information collected. Thus, the ethical aspects of research with human beings were respected, according to Resolution no. 466/2012 [27].

5. Results

Descriptions of hospitalizations of respiratory diseases

(Figure 1) illustrates a typical pattern of hospital admissions (morbidity) for respiratory diseases/ pneumonia/ children and adults, the daily averages for the months 2011-2015.

During the studyperiod (January 1, 2011 to December 31, 2015), the number of hospitalizations for respiratory diseases was 609 (314 children and 295 adults, with a mean of 5 daily admissions, with a minimum of 2 and a maximum of According to the data, a seasonal pattern was observed between the rainys eason, the dry season and the transition period, especially in the quarters (April, May, June , July, August, and September), where the peak of hospitalizations corresponds to the dry season, low rainfall, relative humidity and minimum temperatures, and the period of highest burning rates in the state of Mato Grosso do Sul.



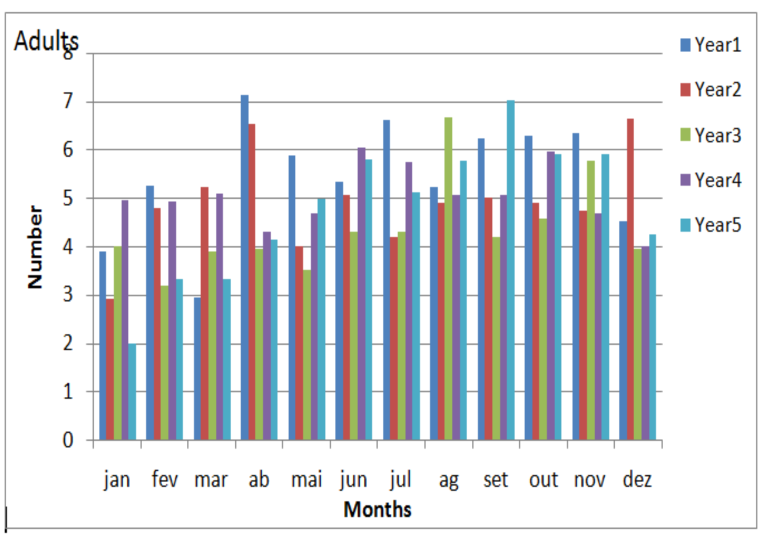


Figure 1: Hospital admissions (morbidity) due to respiratory diseases / pneumonia / in children and adults in Campo Grande, from 2011 to 2015.

5.1. Distributions of Probability and its Estimation of Parameters

The parameters of the estimates of the tested distributions are presented in (Table 3), these parameters are obtained using the ML in the MATLAB software. (Figure 2) shows respectively the histogram of the average hospital admissions of hospitalizations for the years 2011 to 2015 adjusted by the four probability density functions studied and their cumulative frequency adjusted by the four cumulative distribution functions.

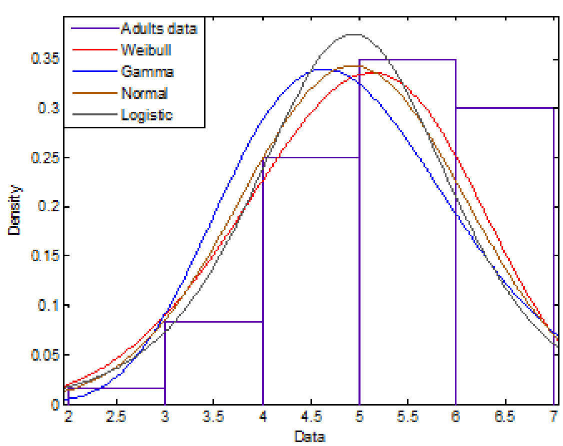
Table 3: Estimated parameters for the distributions studied.

Children

Models	Estimates		R^2	MAE	RMSE	MAPE
Weibull	a= 5.3847	pdf	0,650033856	0,060670537	0,069891164	5,135,100,561
	b= 4.81248	cdf	0,970243009	0,083152694	0,106086726	3,417,287,739
Logistic	mu 4.93266	pdf	0,719799091	0,051104191	**0,039860805**	3,569,588,254
	sigma 0.665318	cdf	0,974237039	0,081689002	**0,062490484**	256,754,589
Gamma	a= 16.7809	pdf	**0,867704992**	**0,03945562**	0,043527192	**2,935,098,395**
	b= 0.293985	cdf	**0,987748319**	**0,080716641**	0,091504676	**2,684,538,084**
Normal	mu 4.93333	pdf	0,73889445	0,052757973	0,060730995	3,709,759,956
	sigma 1.1625	cdf	0,976881289	0,082665248	0,101018504	2,589,206,389

Adults

Models	Estimates		R^2	MAE	RMSE	MAPE
Weibull	a= 5.3847	pdf	0,650033856	0,060670537	0,069891164	5,135,100,561
	b= 4.81248	cdf	0,970243009	0,083152694	0,106086726	3,417,287,739
Logistic	mu 4.93266 sigma 0.665318	pdf	0,719799091	0,051104191	**0,039860805**	3,569,588,254
		cdf	0,974237039	0,081689002	**0,062490484**	256,754,589
Gamma	a= 16.7809	pdf	**0,867704992**	**0,03945562**	0,043527192	**2,935,098,395**
	b= 0.293985	cdf	**0,987748319**	**0,080716641**	0,091504676	**2,684,538,084**
Normal	mu 4.93333	pdf	0,73889445	0,052757973	0,060730995	3,709,759,956
	sigma 1.1625	cdf	0,976881289	0,082665248	0,101018504	2,589,206,389



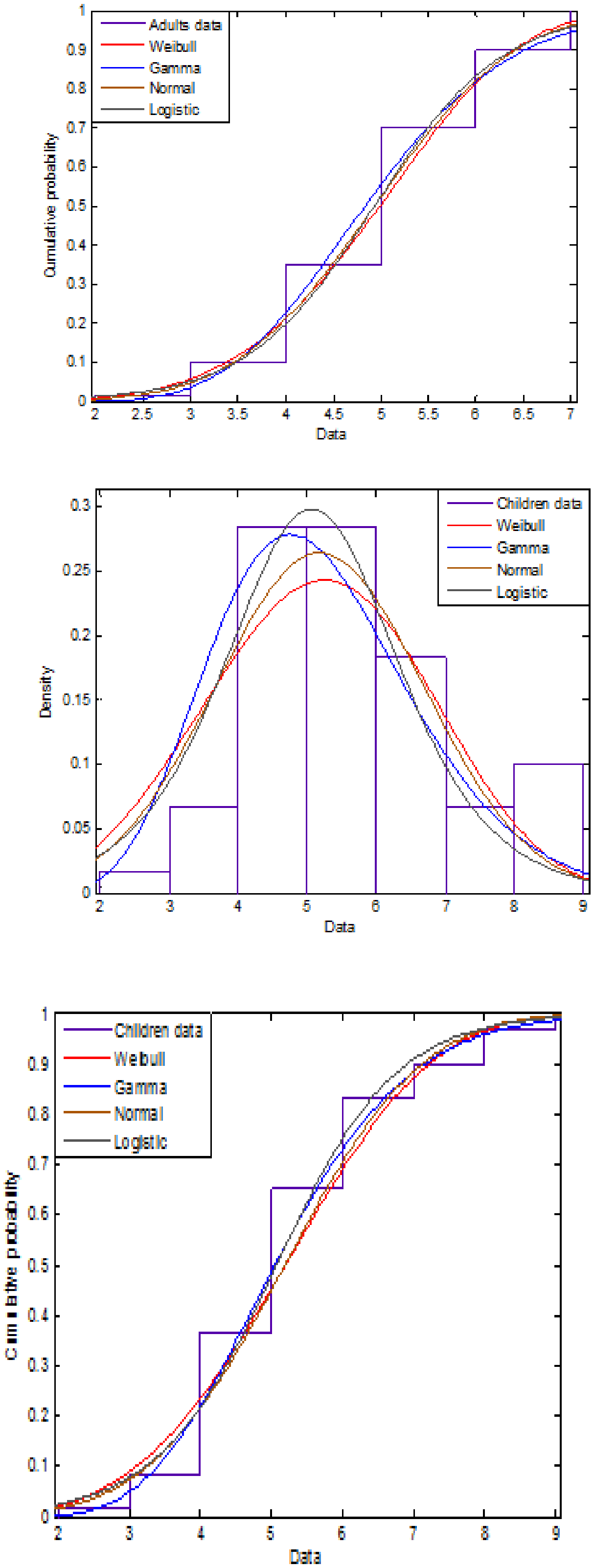


Figure 2: Graphs cdf (left) and pdf (right) of the distributions obtained for the monthly averages of hospital admissions for the years (2011-2015).

5.2. Accuracy Tests

In this work, we consider two statistics associated to cdf (R2 and RMSE) that are sensitive to the central part of the hospital admission distribution and two statistics associated to the pdf that are sensitive to the hospital admission distribution tails [24].

(Table 3) shows the results of suitability tests used. Based on the test the Gamma function is clearly the best distribution, followed by Logistics that best fit the hospital admission data.

To reduce uncertainties in hospital admission estimates, this study compared four candidate distributions (Weibull, Normal, Gamma, Logistics) in order to select the pdf that best matches hospital admissions data. For this, the monthly data of hospital admissions from 2011 to 2015 in the city of Campo Grande, Brazil, are adjusted by the distributions considered. To determine the effectiveness of the statistical models, the performance of four fit quality tests (coefficient of determination, mean square error, mean absolute error and mean absolute error) are performed.

The analysis of the monthly values indicates that the distribution Gamma is the best distribution, it is ranked first. Based on RMSE, Gamma provides the best results. According to the MAE, Gama shows again the best adjustments.

Based on the results obtained, the conclusion we can draw is that the Gamma distribution accurately performs the data. This distribution can be used as an alternative distribution that adequately describes hospital admission data considered in Campo Grande.

References

1. Catalunha MJ, Sediyama GC, Leal BG, Soares CPB, Ribeiro A. Aplicação de cinco funções densidade de probabilidade a séries de precipitação pluvial no Estado de Minas Gerais. Revista Brasileira de Agrometeorologia. 2002; 10: 153-62.
2. Dallacort R, Martins JA, Inoue MH, Freitas PSL, Coletti AJ. Distribuição das chuvas no município de Tangará da Serra, médio norte do Estado de Mato Grosso, Brasil. Acta Scientiarum. Agronomy. 2011; 33: 193-200.
3. Hartmann M, Moala FA, Mendonça MA. Estudo das precipitações máximas anuais em presidente prudente. Revista Brasileira de Meteorologia. 2011; 26: 561-8.
4. Rodrigues JA, Silva APCM, Santos Filho J. Uso de distribuições de probabilidade na modelagem de intensidade de secas ocorridas em Laranjeiras do Sul, PR. Revista Brasileira de Biometria. 2014; 32: 570-83.

5. Kist A, Virgens Filho JS. Análise probabilística da distribuição de dados diários de chuva no estado do Paraná. Revista Ambiente Água. 2015; 10: 172-81.
6. Assis JP, Sousa RP, Batista BDO, Linhares PCF. Probabilidade de Chuva em Piracicaba, SP, Através da Distribuição Densidade de Probabilidade Gama. Revista Brasileira de Geografia Física. 2018; 11: 814-25.
7. Araújo EM, Silva IN, Oliveira JB, Cavalcante Junior EG, Almeida BM. Aplicação de seis distribuições de probabilidade a séries de temperatura máxima em Iguatú-Ceará. RevistaCiênciaAgronômica. 2010; 41: 36-45.
8. Souza A, Aristone F, de Oliveira SS, Olaofe ZO, Kodicherla SPK, Velickovic M et al, Modeling Of The Function Of The Ozone Concentration Distribution Of Surface To Urban Areas. European Chemical Bulletin. 2018; 7: 98-105.
9. Souza A, Olaofe ZO, Kodicherla SPK, Ikefuti P. Probability distributions assessment for modeling gas concentration in Campo Grande, MS, Brazil. European Chemical Bulletin. 2018; 6: 569.
10. Mahiyuddin WRW, Sahani M, Aripin R, Latif MT, Thach TQ, Wong CM. Short-term effects of daily air pollution on mortality. Atmos. Environ. 2013; 65: 69e79.
11. Li Y, Henze DK, Jack D, Henderson BH, Kinney PL. Assessing public health burden associated with exposure to ambient black carbon in the United States. Sci. Total Environ. 2016; 539: 515-25.
12. Souza A, Fernandes WA, Pavao H, Lastoria G, Albrez EA. Potential Impacts Of Climate Variability On Respiratory Morbidity In Children, Infants, And Adults. JornalBrasileiro De Pneumologia (Online). 2012; 38: 708-15.
13. Souza A, Andrade FA, Oguntunde P, Velickovic M et al, Climateindicators and the impact on morbidity and mortality of acute respiratory infections. ADVANCED STUDIES IN MEDICAL SCIENCES. 2018; 6: 5-20.
14. Souza A, Aristone F, Andrade FA, Pavao HG et al. The effects of climate on hospitalization for respiratory diseases by age group. Advanced Studies In Medical Sciences. 2018; 6: 21-35.
15. Souza A, Santos DAS, Ikefuti P. Association between climate variables, pollutants, aerosols and hospitalizations due to asthma. O Mundo Da Saúde (Cusc. Impresso). 2017; 41: 359-67.
16. Santos DAS. et al, A relação das variáveis climáticas na prevalência de infecção respiratória aguda em crianças menores de dois anos em Rondonópolis-MT, Brasil. Ciencia&Saude Coletiva. 2017; 22: 3711-22.
17. Braga ALF, Conceição GMS, Pereira LAA, Kishi HS, Pereira JCR, Andrade MF et al. Air pollution and pediatric respiratory hospital admissions in Sao Paulo, Brazil. J. Occup. Environ. Med. 1999; 1: 95-102.
18. Braga ALF, Saldiva PHN, Pereira LAA, Menezes JJC, Conceiça~o GMS, Lin CA et al. Health effects of air pollution exposure on children and adolescents in Sao Paulo, Brazil. Pediatr. Pulmonol. 2001; 31: 106-13.
19. Viegi G, Maio S, Simoni M, Baldacci S, Annesi-Maesano I. The epidemi- ological link between ageing and respiratory diseases. European respiratory monograph: respiratory diseases in the elderly. Eur. Respir. Soc. 2009; 43: 1-17.
20. de Souza A, Aristone F, Fernandes WA, Olaofe Z, Abreu MC, et al. Statistical Behavior of Hospital Admissions for Respiratory Diseases by Probability Distribution Functions. J InfectDisEpidemiol. 2019; 5: 098.
21. Souza A, Granja SC, Oliveira SS. Estimativa dos parâmetros "C" e "K" do modelo de Weibull e da direção dos ventos para Campo Grande e Dourados, MS, Brasil. RevistaBrasileira de Agrometeorologia, Santa Maria. 1997; 5: 109-14.
22. Tizgui I, El Guezar F, Bouzahir H, Benaid B. Comparison of Methods in Estimating Weibull Parameters for Wind Energy Applications. International Journal of Energy Sector Management 2017;11(4): 650-663.
23. Mohammadi K, Alavi O, McGowan J G. Use of Birnbaum-Saunders distribution for estimating wind speed and wind power probability distributions: A review. Energy Convers. Manag. 2017; 143:109-22.
24. Ouarda, TBMJ, Charron C, Chebana F. Review of criteria for the selection of probability distributions for wind speed data and introduction of the moment and L-moment ratio diagram methods, with a case study. Energy Conversion and Management 2016; 124: 247-65.
25. Jung C, Schindler D. Global comparison of the goodness-of-fit of wind speed distributions. Energy Conversion and Management 2017; 133: 216-34.
26. Hu Q, Wang Y, Xie Z, Zhu P, Yu D. On estimating uncertainty of wind energy with mixture of distributions. Energy. 2016; 112: 935-62.
27. BRASIL. Conselho Nacional de Saúde. Portaria n. 466/2012, de outubro de 2012. Dispõe sobre diretrizes e normas regulamentadoras de pesquisa com seres humanos. Diário Oficial da União, Brasília, 2013; DF, 13 1- 59.