Open Journal of Epidemiology, 2018, 8, 259-275
http://www.scirp.org/journal/ojepi
ISSN Online: 2165-7467
ISSN Print: 2165-7459

Visualizing Fit between Dengue and Climatic Variables on Capitals of the Brazilian Northeast Region by Generalized Additive Models

Julio Cesar Barreto da Silva[1], Hugo Abi Karam[2], Carlos José Saldanha Machado[3]

[1]Programa de Pós-graduação em Meio Ambiente (PPGMA), Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, Brasil
[2]Departamento de Meteorologia do Instituto de Geociências (IGEO), Centro de Ciências Matemáticas e da Natureza (CCMN), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brasil
[3]Instituto de Comunicação e Informação Científica e Tecnológica em Saúde Laboratório de Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz (FIOCRUZ), Ministério da Saúde, Rio de Janeiro, Brasil
Email: barretojcs@gmail.com, hugo@igeo.ufrj.br, saldanha@fiocruz.br

How to cite this paper: da Silva, J.C.B., Karam, H.A. and Machado, C.J.S. (2018) Visualizing Fit between Dengue and Climatic Variables on Capitals of the Brazilian Northeast Region by Generalized Additive Models. *Open Journal of Epidemiology*, **8**, 259-275.
https://doi.org/10.4236/ojepi.2018.84020

Received: October 27, 2018
Accepted: November 26, 2018
Published: November 29, 2018

Abstract

Recent analysis indicates that the numbers of dengue cases may be as high as 400 million/year in the world. According to the Ministry of Brazilian Health, in 2015, there were 1,621,797 probable cases of dengue in the country including all classifications except discarded, the highest number recorded in the historical series since 1990. Many studies have found associations between climatic conditions and dengue transmission, especially using generalized models. In this study, Generalized Additive Models (GAM) was used associated to visreg package to understand the effect of climatic variables on capitals of Northeast Brazilian, from 2001 to 2012. From 12 climatic variables, it was verified that the relative humidity was the one that obtained the highest correlation to dengue. Afterwards, GAM associated with visreg was applied to understand the effects between them. Relative humidity explains the dengue incidence at an adjusted rate of 78.0% (in São Luis-MA) and 82.3% (in Teresina-PI) delayed in, respectively, −1 and −2 months.

Keywords

Aedes aegypti, Dengue, GAM, Visreg Package

1. Introduction

Dengue Fever is fast emerging pandemic-prone viral disease in many parts of the world. Dengue flourishes in urban poor areas, suburbs and the countryside but also affects more affluent neighbor hoods in tropical and subtropical countries. Dengue is a mosquito-borne viral infection causing a severe flu-like illness and, sometimes causing a potentially lethal complication called severe dengue. The incidence of dengue has increased 30-fold over the last 50 years. Up to 50 - 100 million infections are now estimated to occur annually in over 100 endemic countries, putting almost half of the world's population at risk [1].

Tropical countries are the most heavily affected due to environmental, climatic, and social conditions. Studies of climatic variables can improve knowledge and prediction of epidemic seasonality. The climate is an important factor in the temporal and spatial distribution of vector-transmitted diseases as dengue fever [2].

Many works sought to identify climatic influences on dengue, and to evaluate the ability of the climate-based dengue models to describe associations between climate and dengue, simulate outbreaks by generalized additive models—GAMs [3]. This model provides a flexible method for identifying nonlinear covariate effects in exponential family models and other likelihood-based regression models. For this, it used a degree of freedom estimate to assess the importance of covariates based on the expected decrease in the deviance due to smoothing, computable from the trace of the appropriate smoother matrix [4].

We assessed the potential contribution of climatic variables on Dengue Fever (DF) incidences based in GAM, according Hastie and Tibshirani (1990) [5], and we provided suggestions to improve their performance generated from the statistical analyses of the direct and indirect associations.

Mordecai *et al.* (2017) [6] used generalized models associated with R package visreg to understand the impact of temperature on transmission of Zika, dengue and chikungunya. Specifically, Oliveira (2016) [7] used visreg associated with GAM to understand the effect of temperature on ovulation by *Aedes aegypti in* Rio de Janeiro. It wasn't found in the literature, to date, a study involving GAM and visreg package associated with relative humidity.

Ferreira *et al.* (2017) [8] used (Logistic Regression and) GAM associated to Binomial Negative distribution and model offset to understand DF cases relationship meteorological variables, specifically, temperature, rainfall and humidity.

This work aims to identify the risk of DF incidence by the occurrence limits parametrization of climatic variables as a function of the time (months and years), in capitals of the NEB, from January 2001 to December 2012, as from visualizing the fit of regression models arising from of GAM, assuming Poisson Distribution, by cross-sectional plots using two-dimensional contour, by "visreg" package function.

2. Methods

To understand the risk of DF incidence by the occurrence limits parametrization of climatic variables on capitals of the NEB, we conducted the GAM analysis by average monthly data observed from 9 capitals of Brazilian Northeast (NEB), in the period of January 2001 to December 2012. These capitals and their respective codes of the Federative Unit (referring to their States): Aracaju-SE, Fortaleza-CE, João Pessoa-PB, Maceió-AL, Natal-RN, Recife-PE, Salvador-BA, São Luís-MA and Teresina-PI, according **Figure 1**. The data were provided by:

1) Climatological variables: rainfall, in mm, (PRP); minimum, average and maximum temperature, in ˚C, (respectively, T-min, T-mean and T-max); relative humidity, in % (RH); all collected by Meteorological Databank for Education and Research—BDMEP[1], from the National Institute of Meteorology—INMET. From these variables collected, we calculated:

a) Vapour pressure deficit (VPD) and saturated vapour pressure deficit (SVPD), all according Allen *et al.* (1998) [9];

b) Evapotranspiration of Reference (ETO), according Thornthwaite (1948) [10];

c) Annual and monthly heat index, respectively, HI-a and HI-m; as well as, its Function (HI-f), both according to Steadman (1979) [11]; and

d) Human comfort index (HCI), according to Rosenberg (1983) [12].

2) DF cases collected by the site SINAN-Net[2], from the Departamento de Informática do Sistema Único de Saúde—DATASUS; and transformed into DF Incidence Rate; and

3) Annual population size for each of the nine capitals studied, collected by the Sistema de Recuperação Automática—SIDRA[3], from the Brazilian Institute of Geography and Statistics—IBGE.

The monthly reporting DF cases were converted to DF incidence rates, which, according to the Ministry of Health [13], this is defined as the number of confirmed cases (classic and hemorrhagic DF), by 100,000 people in certain geographic space and the current year, and calculated according to Equation (1):

$$\text{DF incidence}=\frac{\text{number of confirmed dengue cases in residents}}{\text{Total resident population in the given period}}\times 100,000 \quad (1)$$

DF incidence are classified by occurrence bands, as criteria of the National Program for Dengue Control—PNCD/MS [13], which considers: 1) low incidence = 0| … 100; 2) average incidence = 100| … |300; and 3) high incidence = 300 … ∞.

All analyses were conducted using the R-Project Software, Version 3.0.3[4].

2.1. Generalized Additive Model

The class of models known as generalized linear models, or GLMs, was formally introduced by Nelder and Wedderburn (1972) [15]. Considering the DF incidences (Y) a response random variable or mean dependent variable, and the

[1]URL: <http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>.
[2]URL: <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinannet/dengue/bases/denguebrnet.def>.
[3]URL: <http://www.sidra.ibge.gov.br/bda/popul/>.
[4]URL: <https://cran.r-project.org/bin/windows/base/old/3.0.3/>.

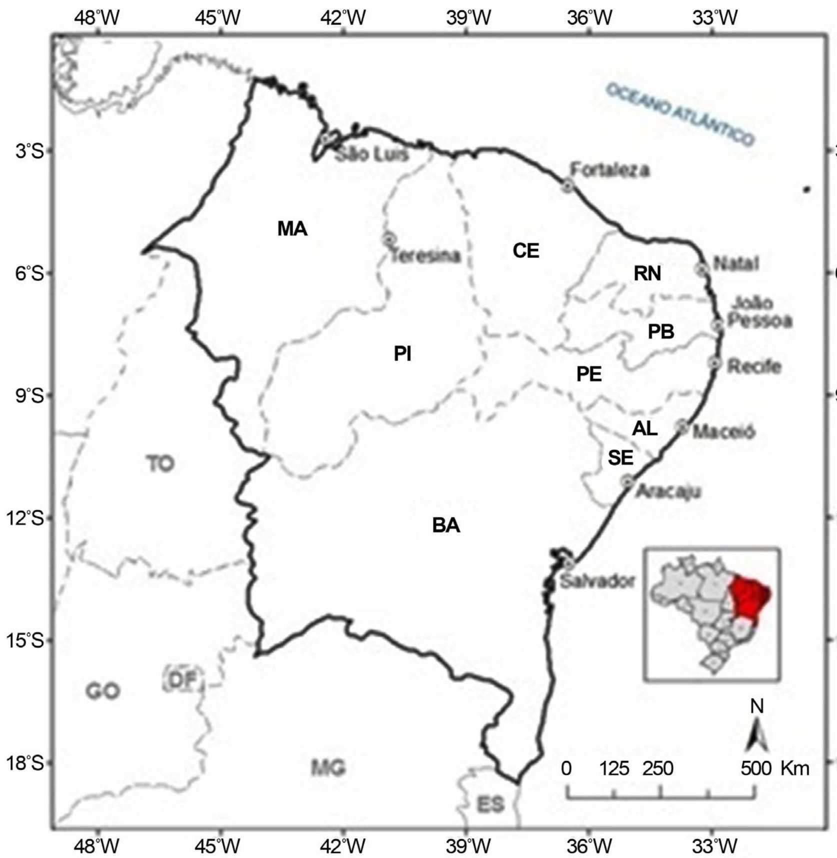


Figure 1. Geographical and political map of the NEB region. Cartographic base: IBGE [14].

temporal (months and years of the time series data) and climatic variables ($X_1, X_2, \ldots, X_p$) a set of predictors or independent/explanatory variables, a regression procedure can be viewed as a method for estimating the expected value of Y given the values of X_i. The standard linear regression model assumes a linear form for the dependency, according Hair Jr. *et al.* (2005) [16], described as:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon \quad (2)$$

where $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$. Given a sample, estimates of $\beta_0, \beta_1, \ldots \beta_p$ are usually obtained by the least squares method.

According Hastie and Tibshirani (1990) [5], GAM consist of a random component, an additive component, and a link function relating that two components, like generalized linear models (GLM). The response Y, the random component, is assumed to have exponential family density:

$$\int Y(y;\theta,\varnothing) = \exp\left\{\frac{y(\theta) - b(\theta)}{a(\varnothing)} + c(y,\varnothing)\right\} \quad (3)$$

where θ is called the natural parameter and $\varnothing$ is the scale parameter. The conditional mean μ of the response variable x to the linear predictor η is related to the set of covariates X_i by a link function g. The quantity:

$$\eta\left(x\right)=s_0+\sum_{i=1}^{p}f_i\left(X_i\right) \tag{4}$$

defines the additive component, where f_i are smooth functions, and the relationship between the conditional mean $\mu\left(x\right)$ and the linear predict $\eta\left(x\right)$ is defined by $g\left(\mu\right)=\eta$. The most commonly used link function is the canonical link, for which $\eta=\theta$. Assuming that $\mu(x)$, is the mean of the Poisson distribution, the dependence of $\mu(x)$ and independent variables X_i, the link function for the Poisson model is the log function $g\left(\mu\right)=\log\left(\mu\right)=\eta$. According Hastie and Tibshirani (1986, 1990) [4] [5], the generalized additive model (GAM) fits a response variable Y by a sum of smooth functions of the explanatory variables, X_i *for* $i = 1, \ldots, p$ by modeling the dependency as:

$$E\left(Y\right)=\beta_0+f_1\left(X_1\right)+f_2\left(X_2\right)+\ldots+f_p\left(X_p\right)+\varepsilon \tag{5}$$

where f_i are smooth functions, $E\left(\varepsilon\right)=0$ and $Var\left(\varepsilon\right)=\sigma^2$.

In order to be estimable, the smooth functions f_i have to satisfy standardized conditions such as $E\left(f_i\left(X_i\right)\right)=0$. GAM extends the parametric form of predictors in the linear model to nonparametric forms. Assuming that Y is normally distributed, an additive model is defined as

$$E\left(Y\right)=s_0+\sum_{i=1}^{p}f_i\left(X_i\right) \tag{6}$$

GAM and GLM can be applied in similar situations, but they serve different analytic purposes. GLM emphasizes estimation and inference for the parameters of the model, while GAM focus on non-parametric data, and this is more suitable for exploring the data and visualizing the relationship between dependent and independent variables, considering the estimation of the smoothing terms f_i in GAM, described in Equation (6) [4].

Smoothers

The spatial distribution was modeled using a bi-dimensional smooth function. A smoother is a tool for summarizing the trend of a response measurement Y as a function of one or more predictor measurements $X_i,\ldots,X_p$. An important property of a smoother is its nonparametric nature. It does not assume a rigid form for the dependence of Y on $X_i,\ldots,X_p$, producing an estimate of the trend that is less variable than Y itself, since of penalized least squares method. Each smoother s_i is controlled by a single smoothing parameter, specificity in the model or choose it automatically by the generalized cross validation method [17] [18] [19]. The GAMs used in this work included a set of directly observed covariates and an s spline smothing function, as depicted in the equations below:

$$\operatorname{logit}\left(Y_i\right)=\beta_0+f\left(\sum_{i=0:12}^{t-L}\beta_k x_k\right)+s\left(\text{month}\right)+s\left(\text{year}\right)+e_i \tag{7}$$

where Y_i is the response variable, in this work the Dengue incidence simulated index, β's are the slope coefficients of the model, so $\exp\left(\beta_0\right)$ is the adjusted odds ratio, x_k are the climatic variables at the individual and household levels as factor of the monthly lags in 0 - 12 times; $s\left(\text{month}\right)$ and $s\left(\text{year}\right)$ are s spline smooth function, and e_i are the residuals. All covariates with a p-value ≤

0.001 in the climatic variable univariate analysis were considered with high significance in the model.

2.2. Chi-Squared Statistic

According Zuur *et al.* (2007) [20], this test is used for comparing models in GLM and GAM to analyze if there is no overdispersion. The chi-square test is one of the most popular hypothesis tests. The Chi-squared Statistic is a measure of how similar two categorical probability distributions are to each other. If the two distributions are identical, the chi-squared statistic is 0, if the distributions are very different, some higher number will result.

$$x^2\left(X,Y\right)=\sum\nolimits_{i=1}^{k}\frac{\left(X_i-Y_i\right)^2}{Y_i} \quad (8)$$

2.3. Package "Visreg"

This interface was used in this work for visualize the fit of regression models arising from of GAM, as from constructing surface by cross-sectional plots using two-dimensional contour or perspective plots. In addition to estimates of this relationship, the package also provides pointwise confidence bands and partial residuals to allow assessment of variability as well as outliers and other deviations from modeling assumptions [21] [22]. The contourlines with high relative risk of DF incidence (Dengue RR) presented in the "visreg" plot were identified on the maps and their climatic limits observed in the model parameterization were considered, as areas with high occurrences of dengue rates.

2.4. Pearson's Correlation Coefficient

Pearson correlation coefficient (r) [23] [24] [25] was used for measuring direction and degree of linear association between dengue and climatic variables, by each capital of the Brazilian Northest. According to Bewick *et al.* (2003) [26], r can be given by:

$$r=\frac{\sum\left(x_i-\bar{X}\right)\left(y_i-\bar{Y}\right)}{\sqrt{\sum\left(x_i-\bar{X}\right)^2\sum\left(y_i-\bar{Y}\right)^2}} \quad (9)$$

where Pearson correlation coefficient or product moment correlation coefficient (r) is a measure of shared variance between two variables, x_i and y_i based in their averages $\bar{X}$ and $\bar{Y}$, and their standard deviations S_x and S_y. The sign indicates a positive or negative direction of the correlation, and the value suggests the power of the relationship between the variables, which value r can vary from −1 to +1, indicating a perfect and very strong positive linear relationship (r = +1), a perfect and very strong negative linear relationship (r = −1), or no linear relationship (r = 0) between the variables [26] [27] [28].

3. Results

In **Table 1**, it is observed the Pearson's correlation coefficient (r) with respective

Table 1. Pearson correlation coefficient (*r*) with respective 95% confidence interval (CI) and p-value, between DF cases and climatic variables, by capital of the NEB.

Capital of the NEB	Climatic Variable	Pearson	CI Lower	95% Upper	p-value
São Luis-MA	PRP	0.0740	−0.0907	0.2347	0.3783
Teresina-PI	PRP	0.0844	−0.0803	0.2446	0.3147
Fortaleza-CE	PRP	0.2707	0.1121	0.4158	0.001
Natal-RN	PRP	0.2430	0.0827	0.3911	0.0033
João Pessoa-PB	PRP	0.2811	0.1232	0.4252	<0.001
Recife-PE	PRP	0.1384	−0.0257	0.2953	0.0980
Maceió-AL	PRP	0.2633	0.1043	0.4093	0.0014
Aracaju-SE	PRP	0.2827	0.1249	0.4266	<0.001
Salvador-BA	PRP	−0.0056	−0.169	0.1582	0.9472
São Luis-MA	RH	0.2494	0.0895	0.3968	0.0026
Teresina-PI	RH	0.3196	0.1646	0.4592	<0.001
Fortaleza-CE	RH	0.3392	0.1859	0.4763	<0.001
Natal-RN	RH	0.3681	0.2177	0.5015	<0.001
João Pessoa-PB	RH	0.2683	0.1095	0.4137	0.0011
Recife-PE	RH	0.1009	−0.6377	0.2601	0.2290
Maceió-AL	RH	0.3794	0.2302	0.5112	<0.001
Aracaju-SE	RH	0.1131	−0.0514	0.2717	0.1770
Salvador-BA	RH	−0.0386	−0.2009	0.1258	0.6462
São Luis-MA	T-min	−0.0740	−0.2347	0.0907	0.3780
Teresina-PI	T-min	−0.1512	−0.3072	0.0127	0.0704
Fortaleza-CE	T-min	−0.2012	−0.3531	−0.0389	0.0156
Natal-RN	T-min	−0.3023	−0.4439	−0.1459	<0.001
João Pessoa-PB	T-min	−0.1533	−0.3091	0.0106	0.0667
Recife-PE	T-min	0.1045	−0.0601	0.2636	0.2127
Maceió-AL	T-min	0.1833	0.0204	0.3368	0.0279
Aracaju-SE	T-min	0.0264	−0.1378	0.1891	0.7539
Salvador-BA	T-min	0.1231	−0.0413	0.281	0.1416
São Luis-MA	T-mean	−0.1867	−0.3399	−0.0239	0.025
Teresina-PI	T-mean	−0.3747	−0.5072	−0.2249	<0.001
Fortaleza-CE	T-mean	−0.2634	−0.4093	−0.1043	0.0014
Natal-RN	T-mean	−0.1622	−0.3173	0.0015	0.0522
João Pessoa-PB	T-mean	−0.1426	−0.2992	0.0215	0.0881
Recife-PE	T-mean	0.0504	−0.1141	0.2122	0.5485
Maceió-AL	T-mean	−0.0073	−0.1707	0.1564	0.9306
Aracaju-SE	T-mean	0.0381	−0.1263	0.2004	0.6504
Salvador-BA	T-mean	0.1231	−0.0413	0.2811	0.1414
São Luis-MA	T-max	0.1281	−0.2857	0.0362	0.1259
Teresina-PI	T-max	−0.3757	−0.508	−0.2260	<0.001

Continued

Fortaleza-CE	T-max	−0.2458	−0.3936	−0.0857	0.0030
Natal-RN	T-max	−0.1407	−0.2974	0.0234	0.0926
João Pessoa-PB	T-max	−0.0560	−0.2176	0.1086	0.5048
Recife-PE	T-max	0.0009	−0.1627	0.1644	0.9917
Maceió-AL	T-max	−0.1486	−0.3048	0.0153	0.0754
Aracaju-SE	T-max	0.0132	−0.1507	0.1764	0.8752
Salvador-BA	T-max	0.0807	−0.0839	0.2411	0.3360
São Luis-MA	SVPD	−0.1824	−0.3359	−0.0194	0.0287
Teresina-PI	SVPD	−0.3765	−0.5087	−0.2269	<0.001
Fortaleza-CE	SVPD	−0.2616	−0.4077	−0.1024	0.0015
Natal-RN	SVPD	−0.1607	−0.3160	0.0029	0.0543
João Pessoa-PB	SVPD	−0.1437	−0.3002	0.0204	0.0858
Recife-PE	SVPD	0.0455	−0.1190	0.2075	0.5882
Maceió-AL	SVPD	−0.0114	−0.1747	0.1524	0.8916
Aracaju-SE	SVPD	0.0325	−0.1318	0.1951	0.6989
Salvador-BA	SVPD	0.1258	−0.0386	0.2835	0.1330
São Luis-MA	VPD	0.2140	0.0523	0.3648	0.0100
Teresina-PI	VPD	0.2153	0.0536	0.3660	0.0095
Fortaleza-CE	VPD	0.2094	0.0474	0.3606	0.0118
Natal-RN	VPD	0.0798	−0.0849	0.2403	0.3415
João Pessoa-PB	VPD	0.1018	−0.0629	0.2610	0.2249
Recife-PE	VPD	0.1916	0.0289	0.3444	0.0214
Maceió-AL	VPD	0.3869	0.2384	0.5177	< 0.001
Aracaju-SE	VPD	0.1110	−0.0535	0.2697	0.1853
Salvador-BA	VPD	0.1215	−0.0429	0.2796	0.1467
São Luis-MA	ETO	0.2059	−0.3574	−0.0438	0.0133
Teresina-PI	ETO	−0.3558	−0.4908	−0.2040	< 0.001
Fortaleza-CE	ETO	−0.3511	−0.4897	−0.1989	< 0.001
Natal-RN	ETO	−0.2668	−0.4124	−0.1079	0.0012
João Pessoa-PB	ETO	−0.2293	−0.3786	−0.0682	0.0057
Recife-PE	ETO	−0.0396	−0.2019	0.1248	0.6372
Maceió-AL	ETO	−0.2027	−0.3545	−0.4049	0.0148
Aracaju-SE	ETO	−0.0264	−0.1892	0.1377	0.7532
Salvador-BA	ETO	0.1046	−0.0600	0.2637	0.2122
São Luis-MA	HCI	−0.0459	−0.2079	0.1185	0.5846
Teresina-PI	HCI	−0.1681	−0.3228	−0.0047	0.0440
Fortaleza-CE	HCI	−0.0319	−0.1945	0.1324	0.7042
Natal-RN	HCI	−0.0776	−0.2382	0.0871	0.3551
João Pessoa-PB	HCI	−0.0613	−0.2226	0.1033	0.4657
Recife-PE	HCI	0.1138	−0.0507	0.2723	0.1744

Continued

Maceió-AL	HCI	0.1305	−0.0338	0.2879	0.1191
Aracaju-SE	HCI	0.0644	−0.1003	0.2256	0.4435
Salvador-BA	HCI	0.1259	−0.0385	0.2836	0.1328
São Luis-MA	HI-a	−0.1569	−0.3124	0.0069	0.0604
Teresina-PI	HI-a	−0.3607	−0.4950	−0.2095	<0.001
Fortaleza-CE	HI-a	−0.1980	−0.3502	−0.0355	0.0174
Natal-RN	HI-a	−0.1312	−0.2886	0.0331	0.1171
João Pessoa-PB	HI-a	−0.1218	−0.2798	0.0427	0.1460
Recife-PE	HI-a	0.0717	−0.0930	0.2325	0.3931
Maceió-AL	HI-a	0.0398	−0.1246	0.2020	0.6360
Aracaju-SE	HI-a	0.0438	−0.1207	0.2059	0.6025
Salvador-BA	HI-a	0.1259	−0.0385	0.2836	0.1328
São Luis-MA	HI-m	−0.1880	−0.3410	−0.0252	0.0240
Teresina-PI	HI-m	−0.3757	−0.5081	0.2260	<0.001
Fortaleza-CE	HI-m	−0.2590	−0.4054	−0.0997	0.0017
Natal-RN	HI-m	−0.1547	−0.3104	0.0091	0.0642
João Pessoa-PB	HI-m	−0.1466	−0.3029	0.0174	0.0796
Recife-PE	HI-m	0.0526	−0.1119	0.2143	0.5311
Maceió-AL	HI-m	−0.0125	−0.1757	0.1514	0.8822
Aracaju-SE	HI-m	0.0333	−0.1310	0.1958	0.6918
Salvador-BA	HI-m	0.1228	−0.0416	0.2808	0.1425
São Luis-MA	HI-f	0.2959	0.1390	0.4382	<0.001
Teresina-PI	HI-f	−0.0424	−0.2046	0.1220	0.6135
Fortaleza-CE	HI-f	0.0977	−0.0670	0.2571	0.2442
Natal-RN	HI-f	−0.3653	−0.4990	−0.2145	<0.001
João Pessoa-PB	HI-f	0.3112	0.1556	0.4518	<0.001
Recife-PE	HI-f	0.0500	−0.1146	0.2119	0.5518
Maceió-AL	HI-f	0.3524	0.2004	0.4879	<0.001
Aracaju-SE	HI-f	−0.0489	−0.2108	0.1156	0.5604
Salvador-BA	HI-f	−0.0512	−0.2130	0.1133	0.5421

95% confidence interval (CI) and p-value, between DF cases and 12 climatic variables, on capital of the NEB. The relative humidity presents the best correlation with DF cases for capitals analyzed, at an absolute average rate of 24.18%, with high significance (p-value < 0.001) observed in four capitals each one. Low correlation is observed with Human Comfort Index and that DF cases, at an absolute rate of 9.1%. In Teresina-PI, there are the best correlations compared to the other capitals tested, at an absolute average rate of 26.67%, and high significance (p-value < 0.001) observed to seven of 12 climatic variables in relationship DF cases. Already, in Aracaju-SE, Recife-PE and Salvador-BA, there are the lower absolute mean correlations and respective no significance p-value observed, suggesting that there are other factors involved in the increase of their DF cases.

Table 2 presents the parametric coefficients of GAM between DF incidence and relative humidity over the period of one year (0 - 12 time-lags), using temporal variables (months) in Teresina-PI and São Luis-MA cities. In Teresina-PI,

Table 2. Parametric coefficients of GAM between DF incidence and relative humidity over the period of one year (0 - 12 time-lags), using temporal variables (months) with *s* term splines smooth, in Teresina-PI and São Luís-MA, in the period from 2001 to 2012.

	Variable	Estimate	SE	z	Pr(>|z|)	Sig
Teresina-PI	(Intercept)	8.375	1. 659	5.047	<0.001	***
Teresina-PI	Lag 0	−0.007	0.007	−0. 920	0. 357	
Teresina-PI	Lag 1	0.029	0.007	3. 993	<0.001	***
Teresina-PI	Lag 2	0.040	0.006	6. 802	<0.001	***
Teresina-PI	Lag 3	0.014	0.005	2. 667	0.008	**
Teresina-PI	Lag 4	−0.044	0.005	−8. 703	<0.001	***
Teresina-PI	Lag 5	−0.051	0.005	−10.068	<0.001	***
Teresina-PI	Lag 6	−0.012	0.005	−2. 465	0.014	*
Teresina-PI	Lag 7	0.000	0.005	−0.085	0. 933	
Teresina-PI	Lag 8	−0.008	0.006	−1. 385	0. 166	
Teresina-PI	Lag 9	−0.033	0.007	−4. 462	< 0.001	***
Teresina-PI	Lag 10	−0.020	0.008	−2. 432	0.015	*
Teresina-PI	Lag 11	−0.012	0.009	−1. 238	0. 216	
Teresina-PI	Lag 12	0.020	0.008	2. 456	0.014	*
São Luís-MA	Intercept	−46.327	8.898	−5.206	<0.001	***
São Luís-MA	Lag 0	0.114	0.022	5.235	<0.001	***
São Luís-MA	Lag 1	0.158	0.020	7.993	<0.001	***
São Luís-MA	Lag 2	0.115	0.016	7.135	<0.001	***
São Luís-MA	Lag 3	0.057	0.014	4.068	<0.001	***
São Luís-MA	Lag 4	0.040	0.015	2.758	0.006	**
São Luís-MA	Lag 5	−0.023	0.015	−1.529	0.126	
São Luís-MA	Lag 6	−0.003	0.018	−0.195	0.846	
São Luís-MA	Lag 7	0.004	0.020	0.203	0.839	
São Luís-MA	Lag 8	0.002	0.020	0.104	0.917	
São Luís-MA	Lag 9	0.021	0.020	1.071	0.284	
São Luís-MA	Lag 10	0.011	0.020	0.566	0.571	
São Luís-MA	Lag 11	0.077	0.019	4.113	<0.001	***
São Luís-MA	Lag 12	0.009	0.017	0.516	0.606	

SE = standard error; z = z-value score; Pr(>|z|) = significance score Z; Sig = significance level: considering "***" when z-value is ≤0.001 (result is "highly significant" with 99.9% of the hypothesis tested being true; that is, the probability (Pr) of the error was less than 0.1%); "**" ≤0.01 (99% of the hypothesis tested is true); and "*" ≤0.1 (9% of the hypothesis tested is true).

GAM shows high significant (p-value < 0.001) association between DF incidence and relative humidity over a range of time-lags 0 - 2, 4 - 5 and 9, being the lag 2 the most significant, with the largest z-value (z = 6.802). Already, in São Luis-MA, the simulated GAM presents high significant level (p-value < 0.001) association between DF incidence and relative humidity over a range of time-lags 0 - 3 and 11, being the lag 1 the most significant, with the largest z-value (z = 7.993).

Table 3 shows the adjust coefficients for GAMs simulated in lags 0 and 1 with DF Cases (using logarithmic function of the population) and DF Incidences, assuming a Poisson distribution, in Teresina-PI and São Luis-MA, in the period from 2001 to 2012. The largest effective degree freedom (edf) values in DF cases simulations indicate nonlinear data when compared to DF incidences. Already, high values of the mean square error (Chi.sq), also simulated with those cases, characterize the overdispersion data. Although the models with DF cases have better fit of the explained deviance; however, your BIAS are extremely high, making the models with DF incidences more parsimonious and therefore more suitable for use [29]. In Teresina-PI, the modeling by GAM with relative humidity over a time-lag 2 explain 82.3% of the deviance on DF incidences while São Luis-MA over a time-lag 1 explain 78.0% of the deviance on DF incidences, with significant effects in the adjust coefficients with low effective degree freedom, respectively, 6.067 and 7.276; and low estimate of the intercept and respective z-value, making it the best simulated model. In the lag 0 (no lag effect), both models presented the best estimate and z-value, although they had the lowest *R-adjusted* between the variables measured, 0.699 in both.

Figure 2 shows the distribution of DF incidence and relative humidity as

Table 3. Parametric coefficients of GAM between DF and relative humidity, for time-lags (0 - 1 lags in São Luis-MA and 0 - 2 lags in Teresina-PI) using temporal variables (months) with s term splines smooth, simulated with DF cases (with logarithmic function of population) in the period of 2001 and 2012.

			offset					Intercept		
	Model	Dengue	log (pop)	R-sq (adj)	DE (%)	edf	Chi.sq	Estimate	SE	z
Teresina	Lag 0	Case	Yes	0.714	77.3	8.991	22949.0	4.1331	0.0036	1159.0
Teresina	Lag 0	Incidence	No	0.699	76.3	7.887	154.9	1.8642	0.0409	45.6
Teresina	Lag 2	Case	Yes	0.750	79.3	8.989	30581.0	4.1376	0.0036	1162.0
Teresina	Lag 2	Incidence	No	0.754	82.3	6.067	118.2	2.5199	0.0356	70.81
São Luís	Lag 0	Case	Yes	0.714	77.3	8.991	22949.0	4.1331	0.0036	1159.0
São Luís	Lag 0	Incidence	No	0.699	76.3	7.887	154.9	1.8642	0.0409	45.6
São Luís	Lag 1	Case	Yes	0.750	79.3	8.989	30581.0	4.1376	0.0036	1162.0
São Luís	Lag 1	Incidence	No	0.726	78.0	7.276	210.2	1.8726	0.0408	45.9

R-sq = R square adjusted; DE = explained deviance; edf = effective degree freedom, chi.sq = quadratic mean error; SE = standard error; z = z-value score.

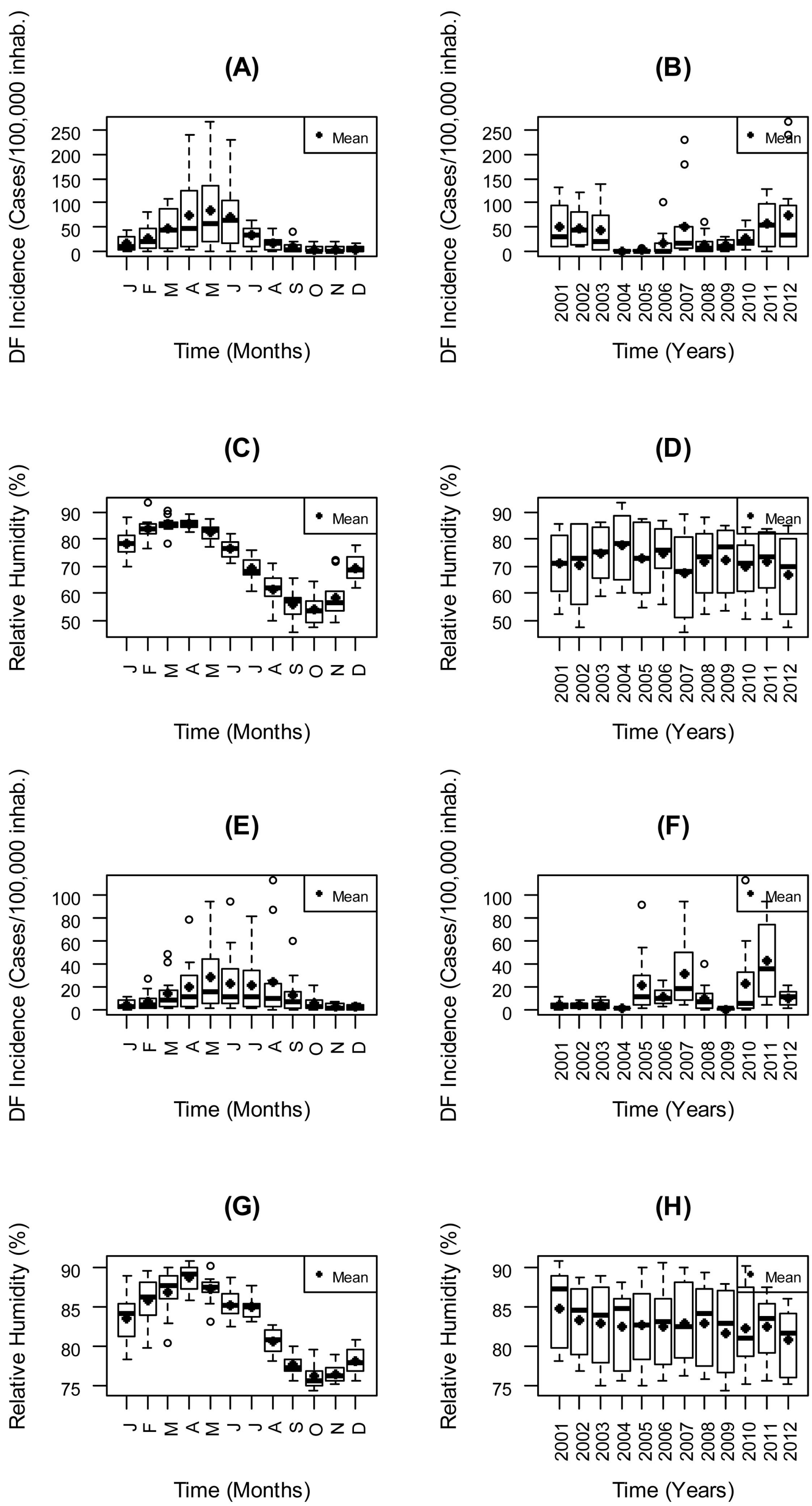


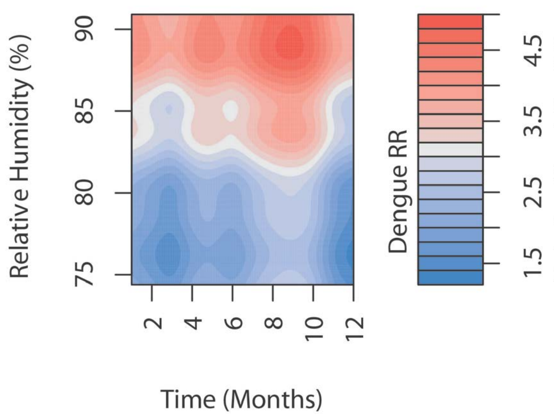
Figure 2. Seasonality effect, by boxplot, on DF incidence (DF Cases/100.000 inhabitants) in function of months (A and E) and years (B and F); and relative humidity (%) in function of months (C and G) and years (D and H), in Teresina-PI and São Luis-MA, respectively, from January 2001 to December 2012.

time function, in Teresina-PI and São Luis-MA. The seasonal trend of that incidence over monthly and annual time-frequency was observed.

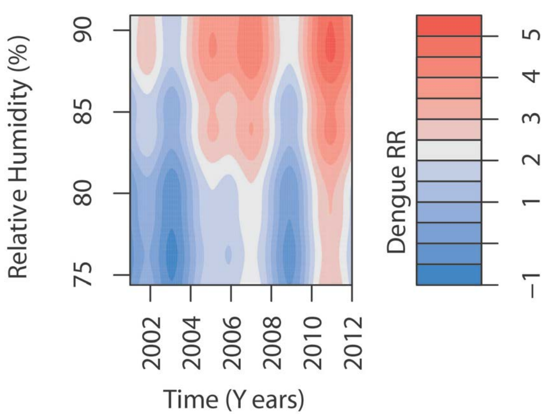
In relation to Teresina-PI, Figure 2(A) shows an increase in the DF incidence from the month of January with peak and highest mean in May, that coincides with the lagged relative humidity in 2 months or March (Figure 2(C)); declining from this month with lower rates in December, that coincides with the lagged relative humidity in 2 months or October (Figure 2(C)). It is also observed three annual periods for the occurrence of the DF epidemiological cycles, with peak in 2010-2011, 2007 and 2001-2003, according Figure 2(B).

In relation to São Luis-MA, Figure 2(E) shows an increase in the DF incidence from the month of January with peak and highest mean in May that coincides with the lagged relative humidity in 1 month, according to Figure 2(G), declining from this month with lower rates in November and December. The highest occurrence and average of DF incidences were recorded in the years 2011, 2007, 2010 and 2005, in this descending order, Figure 2(F).

Figure 3 shows the visualization of the effect of simultaneous variance between relative humidity and time (months and years, Figure 3(A) and Figure 3(B), respectively), in relationship to DF incidence risk (Dengue RR), simulated on DF incidences, from January 2000 to December 2012, by "visreg" function on simulated regression GAM using penalized *s* splines smoothing, in São Luís-MA. Figure 3(A) shows a large nucleus limited to between 87.0% and 90.0% relative humidity between August and October months, with high relative risk Dengue (RR = 5.0), that is, high DF incidences. Comparing this figure to contour in function of the years, Figure 3(B), 3 nuclei are observed, characterizing the years of highest DF incidences, being 83.0% and 90.0% of relative humidity range highly significant to occurrences those incidences (RR ≥ 4.0).



(A)



(B)

Figure 3. Visualization of the relationship between relative humidity on lag 1 and temporal variables (months and years), as response to Dengue Relative Risk or Dengue RR (simulated on DF incidences) by GAM regression with Poisson distribution using penalized *s* splines smoothing, in São Luis-MA, from January 2000 to December 2012. The legend presents in a gradual degree of Dengue RR, which ranges from −1 to 5 in Figure 3(B), with 5 being the positive chance of having incidence of dengue in the studied population increased by 5x, while −1 would be the possibility of RR in −1x.

4. Discussion

All numerical output of GAM and respective intercepts in the capitals of the NEB have obtained setting on p-value of 0.001. The capital of João Pessoa-PB is the one with the smaller values of mean squared error (Chi.sq); in other words, the values of the estimated parameters of climatic variables around the true value of DF cases present a greater accuracy and precision in the quality of response by GAM. According Bolker (2008) [29], the mean square error represented by the sum of the variance and bias square indicates the quality of an estimator and shows the total change around a true value, in this study, the DF cases.

We found a high correlation of DF incidence with relative humidity is lagged in 1 and 2 months, respectively, in São Luis and Teresina cities. Wu *et al.* (2007) [30] observed cross-correlation with statistical significance between DF incidence and relative humidity over a range of time-lags from −1 to −3 months, above all the most dominant effect at a lag −2 months ($r = 0.202$, $p < 0.005$).

Ehelepola and Ariyaratne (2016) [31] evidenced in their study a median increase in 7x of dengue incidences, for a relative humidity of 86%, according to figure 6 of that study. Neto and Rebêlo (2004) [32], studying the association between dengue cases and climatic variables in São Luis-MA, from 1997 to 2002, verified that dengue cases are directly related to the increase of precipitation and relative humidity, with a positive correlation this variable of 76.0% ($r = 0.76$; $p < 0.05$). In addition, the authors identified peaks of relative humidity in the months of March and April, an average variation from 85.6% (March 1999) to 89.3% (April 1997), while the highest percentage of dengue was presented in May, with an average percentage of 20.2% of cases recorded. While in this study, we identify an *r*-adjusted between these variables of 0.75 for a −1 month lag of relative humidity in relation to dengue.

5. Conclusions

The formulation of GAM model is nearly exactly the same as for GLM. These models use all the same families and link functions; but GAM is wrapping the predictors in a non-parametric smoother function, in this paper, specifically, the *s* spline. The GAM fit is more sensitive to minimizing deviance (higher wiggliness) than the default fit of the loess function. This model is also able to minimize deviance based on the logit transformation. The model output shows that an overall (parametric) intercept is fitted (the mean) on the scale of the logit transformation (logarithmic population of the capitals studied).

Modeling by GAM, assuming a Poisson distribution, explained 82.3% of the deviance of DF incidences, and significant effects were found in the estimates of all climatic variables on dengue; however, the high values of the effective degrees of freedom (edf) of smooth functions indicate that the association between dengue and climate is highly nonlinear. The estimate initially found, by the GLM and GEEGLM models for these studied variables, was too high, indicating the overdispersion data, however regressions by GAM reduced significantly excess dispersion presented in the proportion of deviations from the response shown in simulations by GLM and GEEGLM, *i.e.*, not shown here. Our results were robust to other model specifications with different controls for long-term and seasonal trends. It is suggested that the models proposed in this paper are used by surveillance agencies for planning, prevention and control of Dengue Incidence.

From 12 climatic variables, it was verified that the relative humidity was the one that obtained the highest correlation to dengue in six of nine capitals of the NEB, with high significance ($p < 0.001$) in Teresina-PI, Fortaleza-CE, Natal-RN and Maceió-AL. Afterwards, GAM associated with visreg was applied to understand the effects between them. March and April months show the sensibility of the use of GAM for the analysis of that correlation. Relative humidity explains the dengue at an adjusted rate of 78.0% (in São Luis-MA) and 82.3% (in Teresina-PI) delayed in, respectively, −1 and −2 months.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] World Health Organization (2015) Dengue Topics. http://www.who.int/topics/dengue/en/

[2] Rosa-Freitas, M.G., Schreiber, K.V., Tsouris, P., Weimann, E.T.S. and Luitgards-Moura, J.F. (2006) Associations between Dengue and Combinations of Weather Factors in a City in the Brazilian Amazon. *Revista Panamericana de Salud Pública*, **20**, 256-267. https://doi.org/10.1590/S1020-49892006000900006

[3] Honório, N.A., Nogueira, R.M.R., Codeço, C.T., Carvalho, M.S., Cruz, O.G., *et al.* (2009) Spatial Evaluation and Modeling of Dengue Seroprevalence and Vector Density in Rio de Janeiro, Brazil. *PLOS Neglected Tropical Diseases*, **3**, e545. https://doi.org/10.1371/journal.pntd.0000545

[4] Hastie, T. and Tibshirani, R. (1986) Generalized Additive Models. *Statistical Science*, **3**, 297-318. https://doi.org/10.1214/ss/1177013604

[5] Hastie, T.J. and Tibshirani, R.J. (1990) Generalized Additive Models. Chapman and Hall/CRC, London.

[6] Mordecai, E. A.; Cohen, J. M.; Evans, M. V.; Gudapati, P.; Johnson, L. R. *et al.* (2017) Detecting the Impact of Temperature on Transmission of Zika, Dengue, and Chikungunya Using Mechanistic Models. *PLOS Neglected Tropical Diseases*, **11**, 1-18. https://doi.org/10.1371/journal.pntd.0005568

[7] Oliveira, S. de S. (2016) Análise espacial e temporal da infestação por *Aedes aegypti* mensurada por ovitrampas para geração de alerta precoce de dengue no município do Rio de Janeiro. Dissertação de Mestrado. Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública Sergio Arouca, Rio de Janeiro, 136 p.

[8] Ferreira, D.A. da C., Degener, C.M., Marques-Toledo, C. de A., Bendati, M.M., Fetzer, L.O., *et al.* (2017) Meteorological Variables and Mosquito Monitoring Are Good Predictors for Infestation Trends of *Aedes aegypti*, the Vector of Dengue, Chicungunya and Zika. *Parasites & Vectors*, **10**, 78. https://doi.org/10.1186/s13071-017-2025-8

[9] Allen, R.G., Pereira, L.S., Raes, D. and Smith, M. (1998) Crop Evapotranspiration (Guidelines for Computing Crop Water Requirements). *Fao Irrigation and Drainege*, **56**, 297.

[10] Thornthwaite, W.C. (1948) An Approach toward a Rational Classification of Climate. *Geographical Review*, **38**, 55-94. https://doi.org/10.2307/210739

[11] Steadman, R.G. (1979) The Assessment of Sultriness: Part I: A Temperature-Humidity Index Based on Human Physiology and Clothing Science. *Journal of Applied Meteorology*, **18**, 861-884. https://doi.org/10.1175/1520-0450(1979)018<0861:TAOSPI>2.0.CO;2

[12] Rosenberg, N.J., Bland, B.L. and Verma, S.B. (1983) Microclimate: The Biological Environment. John Wiley & Sons, New York, 467 p.

[13] Ministério da Saúde da República Federativa do Brasil—MS/Brasil. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica (2009) Diretrizes nacionais para prevenção e controle de epidemias de dengue. Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância Epidemiológica. Ministério da Saúde, Brasília, 160 p. (Série A. Normas e Manuais Técnicos). http://bvsms.saude.gov.br/bvs/publicacoes/diretrizes_nacionais_prevencao_controle_dengue.pdf

[14] Instituto Brasileiro de Geografia e Estatística—IBGE. Ministério do Planejamento, Orçamento e Gestão da República Federativa do Brasil—MP/Brasil (2010) Atlas geográfico escolar: ensino fundamental do 6º ao 9º. Ministério do Planejamento, Orçamento e Gestão, Instituto Brasileiro de Geografia e Estatística—IBGE. IBGE. Rio de Janeiro. 168 p/ il. color. https://biblioteca.ibge.gov.br/visualizacao/livros/liv49956_capa_apres_sum.pdf

[15] Nelder, J. and Wedderburn, R. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society A*, **135**, 370-384. https://doi.org/10.2307/2344614

[16] Hair Jr., J.F., *et al.* (2005) Análise multivariada de dados. Bookman, São Paulo.

[17] Craven, P. and Wahba, G. (1979) Smoothing Noisy Data with Spline Functions. *Numerical Mathematics*, **31**, 377-403. https://doi.org/10.1007/BF01404567

[18] Wahba, G. (1990) Spline Models for Observational Data. Society for Industrial and Applied Mathematics, Philadelphia. https://doi.org/10.1137/1.9781611970128

[19] Wood, S.N. (2006) Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC, London. https://doi.org/10.1201/9781420010404

[20] Zuur, A., Ieno, E.N. and Smith, G.M. (2007) Analyzing Ecological Data Statistics for Biology and Health. Springer Science & Business Media, Berlin, 672 p. https://doi.org/10.1007/978-0-387-45972-1

[21] Breheny, P. and Burchett, W. (2017) Visualization of Regression Models: Using Visreg. *The R Journal*, **9**, 56-71. https://journal.r-project.org/archive/2017/RJ-2017-046/index.html

[22] Breheny, P. and Burchett, W. (2018) Visualization of Regression Models. Package "Visreg". Version 2.5-0. https://cran.r-project.org/web/packages/visreg/visreg.pdf

[23] Pearson, K. (1901) On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, **2**, 559-572. https://doi.org/10.1080/14786440109462720

[24] Pearson, K., Fisher, R. and Inman, H.F. (1994) Karl Pearson and R. A. Fisher on Statistical Tests: A 1935 Exchange from Nature. *American Statistician*, **48**, 2-11. https://doi.org/10.2307/2685077

[25] Moore, D.S. (2007) The Basic Practice of Statistics. 4th Edition, Freeman, New York.

[26] Bewick, V., Cheek, L. and Ball, J. (2003) Statistics Review 7: Correlation and Regression. *Critical Care*, **7**, 451-459. https://doi.org/10.1186/cc2401

[27] Figueiredo Filho, D.B., Silva, J.R. and Da, J.A. (2009) Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r). *Revista Política Hoje*, **18**, 115-146. http://www.ufpe.br/politicahoje/index.php/politica/article/view/6/6

[28] Kozak, M. (2009) What Is Strong Correlation? *Teaching Statistics*, **31**, 85-86. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9639.2009.00387.x https://doi.org/10.1111/j.1467-9639.2009.00387.x

[29] Bolker, B.M. (2008) Ecological Models and Data in R. Princeton University Press, Princeton and Oxford, 508 p.

[30] Wu, P.C., Guo, H.R., Lung, S.C., Lin, C.Y. and Su, H.J. (2007) Weather as an Effective Predictor for Occurrence of Dengue Fever in Taiwan. *Acta Tropica*, **103**, 50-57. https://linkinghub.elsevier.com/retrieve/pii/S0001706X07001271 https://doi.org/10.1016/j.actatropica.2007.05.014

[31] Ehelepola, N.D.B. and Ariyaratne, K. (2016) The Correlation between Dengue Incidence and Diurnal Ranges of Temperature of Colombo District, Sri Lanka 2005-2014. *Global Health Action*, **9**. https://doi.org/10.3402/gha.v9.32267

[32] Neto, V.S.G. and Rebêlo, J.M.M. (2004) Epidemiological Characteristics of Dengue in the Municipality of São Luís, Maranhão, Brazil, 1997-2002. *Cadernos de Saúde Pública*, **20**, 1424-1431.