

21 **Abstract**

22 In this work we studied the relationship between dengue incidence in Cali and the climatic
23 variables that are known to have an impact on the mosquito and were available (precipitation,
24 relative humidity, minimum, mean, and maximum temperature). Since the natural processes
25 of the mosquito imply that any changes on climatic variables need some time to be visible
26 on the dengue incidence, a lagged correlation analysis was done in order to choose the
27 predictor variables of count regression models. A Principal Component Analysis was done to
28 reduce dimensionality and study the correlation among the climatic variables. Finally, aiming
29 to predict the monthly dengue incidence, three different regression models were constructed
30 and compared using de Akaike information criterion. The best model was the negative
31 binomial regression model, and the predictor variables were mean temperature with a 3-
32 month lag and mean temperature with a 5-month lag as well as their interaction. The other
33 variables were not significant on the models. And interesting conclusion was that according
34 to the coefficients of the regression model, a 1°C increase in the monthly mean temperature
35 will reflect as a 45% increase in dengue incidence after 3 months. The rises to a 64% increase
36 after 5 months.

37 **Key words**

38 Dengue, Negative binomial regression model, random forest regression, climatic variables.

39 **Author Summary**

40 Dengue is transmitted by the bite of an infected mosquito, and mosquitoes, in turn, are
41 affected by climatic conditions. In this work studied the relationship between dengue
42 incidence in Cali and climatic variables, namely precipitation, relative humidity, minimum

43 temperature, mean temperature, and maximum temperature using statistical methods. Since
44 this is a natural and biological process, the changes in climatic conditions need time to have
45 a visible effect on dengue incidence, hence we identified the significant climatic variables
46 and the time they take to have a visible effect on dengue incidence. Then, we created three
47 different models for predicting dengue incidences using the lagged variables and picked the
48 best one. We concluded that the most critical variable is mean temperature with a 3- and 5-
49 month lag. We also found that a 1°C increase in the monthly mean temperature will reflect
50 as a 45% increase in dengue incidence after 3 months. The rises to a 64% increase after 5
51 months.

52 **1. Introduction**

53 Dengue is the viral disease transmitted by *Aedes aegypti* with the biggest worldwide
54 spreading. It's common in warm and humid areas all around the world. It's transmitted by
55 the biting of infected mosquitoes, and it can cause high fever, nausea, vomiting, rash and pain
56 in different parts of the body (eyes, muscles, joints or bones). When severe, it may also cause
57 internal bleeding and even death [1].

58 Every year around 50 million dengue cases are registered, and approximately 2.500 million
59 people live in countries where this disease is endemic [2]. Furthermore, dengue affects Cali,
60 a city in the southwest of Colombia, chronically. In fact, yearly mean rates of dengue in Cali
61 reach almost 100 cases per 100.000 inhabitants [3].

62 On the other hand, it's well known that climatic conditions play a big part on the reproduction
63 and maturation of *Aedes aegypti*. Warm temperature shortens the incubation time of the
64 larvae, while precipitation directly affects the availability of breeding places [4–6] Also,

65 temperature exerts a significant influence and regulates the development of mosquitoes [7].
66 However, excesses of precipitation might imply a reduction of the population since neither larvae
67 nor pupae survive in these conditions [8–11]. Additionally, El Niño phenomenon (ELSO) is
68 related with dengue incidence and the effects take several months to be reflected on it
69 depending on the region [12–14].

70 Concerning the modeling of dengue, various count regression models have been successfully
71 used [15–18]. Amongst them there is the negative binomial model since it, unlike the Poisson
72 model, doesn't assume equality between the mean and the variance. Using this type of models
73 [19–22] found that temperature and relative humidity are important variables for predicting
74 dengue incidence.

75 In Cali, diverse intervention measures have been implemented. In 2022, the activities carried
76 out include drain inspections, household visits for eliminating hatching, control campaigns,
77 biological control, and fumigation. For 2023, the Public Health Department of Cali will begin
78 its work with initiatives aiming at education and prevention [23]. Nevertheless, it is still
79 crucial to continue the research on the impact of climatic variables on dengue incidence to
80 prevent endemic spikes and adjust control strategies.

81 Because of the way dengue is propagated, focusing control strategies on the vector is
82 essential, which generally imply intensive and expensive efforts. Hence, this control schemes
83 must be implemented strategic fashion in order to maximize their efficiency and minimize
84 costs. Because of this, the study of dengue incidence in relation with climatic variables
85 associated with the reproduction and development of *Aedes aegypti* should not be
86 disregarded.

87 The main goal of this research consisted of analyzing the relationship and delay between
88 climatic variables and dengue incidence. We sought to describe and quantify how the climatic
89 patterns are directly linked with dengue incidence throughout statistical analysis. Data of Cali
90 was selected because this city presents one of the highest dengue infection rates in Colombia.

91 **2. Materials and methods**

92 **2.1. Area of study**

93 Cali, the third biggest city in Colombia, is the capital of the department of Valle del Cauca.
94 It's located on the south of the department, on the valley formed by the western and central
95 branches of the Andes Mountain range, at a mean altitude of 1000 meters above sea level.
96 Cali is also located on the valley of Cauca River, the second most important of Colombia.
97 Cali's geographical coordinates are 3°27'00"N 76°32'00"O. Cali's weather is warm and
98 humid, with an average temperature of 24°C.

99 The weather in Cali es characterized for being warm and dry, classifying as tropical with dry
100 summers according to the Köppen Climate Classification. The western branch of the
101 mountain range plays an important role stopping the humid air currents coming from the
102 Pacific Ocean, although the maritime breezes do indeed reach the city. The mountain range
103 has an average altitude of 2000 meters above sea level at the north of the city, reaching up to
104 4100 meters on the south, where the highest pluviosity occurs. On average, annual
105 precipitation lies from 900 mm on the driest areas up to 1800 on the rainiest, being 1483 mm
106 the average over the majority of Cali's metropolitan area. Mean temperature es 24.0°C
107 (73.6°F). From December to February and July to Augusts it's considered the dry season,
108 while from March to May and September to November is the rainy season [24].

109 **2.2. Data collection**

110 For this study, data of monthly dengue cases provided by the Public Health Department of
111 Cali was used, encompassing a time period from January of 2015 to December 2021.
112 Similarly, from the Hydrology, Meteorology and Environmental Studies Institute (IDEAM)
113 monthly data were collected for the following climatic variables: maximum, minimum and
114 average temperature (°C), precipitation (mm) and relative humidity (%) on the same time
115 period.

116 **2.3. Statistical methods**

117 **2.3.1. Random Forest Regression**

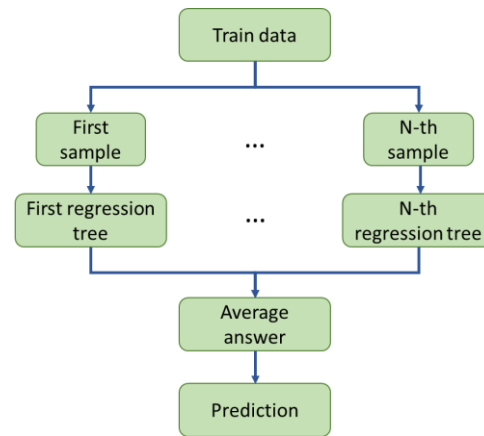
118 The random forest regression technique was used for missing data imputation with python's
119 *sklearn* library on Google Collab. This machine learning method stands out for its efficiency
120 on both classification and regression problems, although in this work we only care for
121 regression problems. This method is based on the construction of a set of decision trees that
122 will be used to predict continuous values. This technique is increasingly been used in diverse
123 areas for the processing of big volumes of data [25,26].

124 As shown on Fig 1, the main idea of random forest regression is to create multiple different
125 regression trees trained with different samples from the training set. Thus, each tree will make
126 different individual predictions that then will be averaged to produce a single final prediction.
127 This method tends to be better than using a single regression tree since the latter is highly
128 sensible to the training data and so, the averaged prediction is more robust.

129

130

Fig. 1 Random Forrest regression algorithm



131

132 Similarly, regression trees are a machine learning technique used for prediction problems. In
133 other words, we want to predict the value of a continuous variable y using a set of predictor
134 variables or covariables X . Hence, a tree is constructed such that on each node examples are
135 split in a way where on each new node the variance of y is the least possible. This implies
136 picking on every node a suitable covariable and splitting point c . This process is repeated
137 until a stop criterion is reached which may depend on the depth of the tree or the number of
138 examples on the nodes.

139 It's essential to remark that each of the regression trees are different, and therefore return
140 different predictions, because they are constructed from different random samples from the
141 data base. This is called bagging and is used when working with noise data sets.

142 **2.3.2. Correlation analysis**

143 Since the effect of climatic variables over the dengue incidence is not immediate, a lagged
144 correlation analysis was made, a common tool of time series analysis, in order to select the
145 variables and corresponding lag to be the predictors on the regression model. Subsequently,

146 the number of dimensions was reduced by doing a principal component analysis (ACP) also
147 avoiding multicollinearity issues.

148 **2.3.3. Principal component analysis**

149 ACP is a useful statistical technique for synthesizing information or reducing the number of
150 variables on a data set without losing important information. During ACP new variables are
151 created called principal components, which are linear combinations of the original variables
152 and, additionally, they are pairwise mutually independent.

153 In this work, ACP was used for two purposes. Firstly, as a data exploration tool in order to
154 study the correlation between the climatic variables and, secondly, to reduce the amount of
155 predictor variables in the statistical model. The similarities between these variables were
156 studied and only one was chosen amongst the groups of homogeneous variables so it acts a
157 representative, since multiple variables with like behavior is not needed. This analysis was
158 made using SPSS.

159 **2.3.4. Regression models**

160 Keeping in mind that dengue incidence is a count variable, three different regression models
161 were considered for this project, namely, Poisson, negative binomial, and Poisson-inverse
162 Gaussian. The Poisson model is one of the more classical ones for count data. It assumes that
163 the data follow a Poisson distribution with the following probability distribution function:

$$164 \quad P(Y_i|\mu_i) = \frac{e^{-\mu_i} \mu_i^{Y_i}}{Y_i!}.$$

165 Here, Y_i is a random variable with mean and variance μ_i . For constructing regression models,
166 μ_i is expressed as a function of the independent variables throughout the natural log as a link
167 function:

$$168 \quad \ln \mu_i = \beta' x_i.$$

169 Similarly, the dependent variable is link, throughout the natural log, whit a linear function of
170 the independent variables:

$$171 \quad \ln Y = \beta_0 + \beta_1 X_1 + \dots + \beta_r X_r .$$

172 The negative binomial model is similar to the Poisson model with the key difference that it
173 relaxes the assumption of equality of mean and variance, which improves it when dealing
174 with over dispersed data (when variance is bigger than the mean). This is regularly the case
175 with real data since they can be zero [27]. Therefore, negative binomial regression may be
176 considered as a generalization of the Poisson regression.

177 For this model the value of μ differs from the one it takes on a Poisson regression, namely,

$$178 \quad \mu_i = \exp(\beta' x_i + e_i) = \exp(\beta' x_i) \exp(e_i)$$

179 were e_i is a heterogeneity parameter and $\exp(e_i) \sim \Gamma(\alpha^{-1}, \alpha^{-1})$. The density function for Y_i
180 is:

$$181 \quad P(Y_i|X_i) = \frac{\Gamma(Y_i + \alpha^{-1})}{\Gamma(Y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{Y_i} + \left(\frac{\mu_i}{\alpha^{-1} + \mu_i} \right).$$

182 Finally, the Poisson-inverse Gaussian (PIG) regression is very similar to the negative
183 binomial, but instead of combining the Poisson distribution with a Gamma distribution, it

184 uses an inverse Gaussian distribution. This model is particularly good for data with a high
185 initial peak and that might be skewed to the far right, or highly over dispersed data [28].

186 For the PIG regression α represents the dispersion parameter and can also be parametrized
187 as $\phi = 1/\alpha$. Thereby, the probability distribution function is:

$$188 \quad P(Y; \mu, \alpha) = \sqrt{\frac{\phi}{2\pi Y^3}} \exp\left(\frac{-\phi(Y - \mu)^2}{2\mu^2 Y}\right)$$

189 with $Y, \mu, \phi > 0$.

190 **3. Results**

191 **3.1. Data imputation**

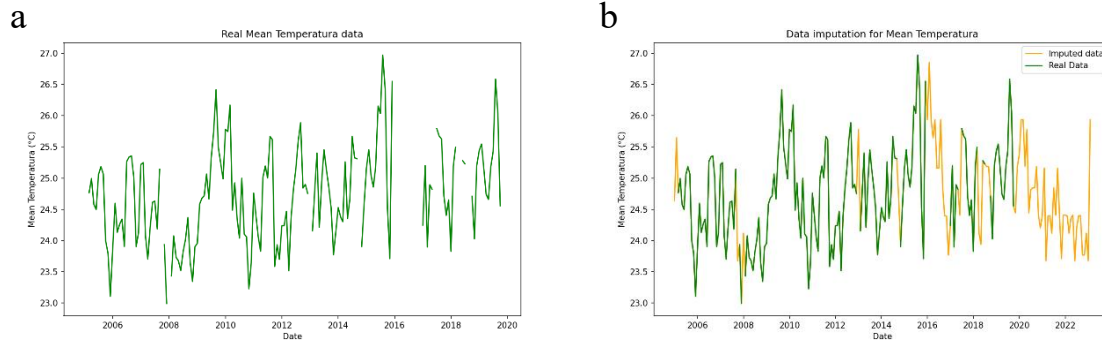
192 The first obstacle to present itself was the presence of missing data on the climatic data set.
193 This a fairly common difficulty when working with climatic data. A process of data
194 imputation was necessary beforehand. For this purpose, the random forest regression method
195 was used [29]. First, for every variable, a second meteorological station near Universidad del
196 Valle without so many missing values was selected (Palmira or Jamundí), and then, the
197 missing values of this new stations were replaced with the mean, and finally, a random forest
198 regression model was trained using the secondary station to predict the missing values of the
199 station at Universidad del Valle. As an example, in Fig 2 shows graphs for the original data
200 and the completed version with imputation for mean temperature.

201

202

203

Fig. 2 a) original data; b) data with imputation



204

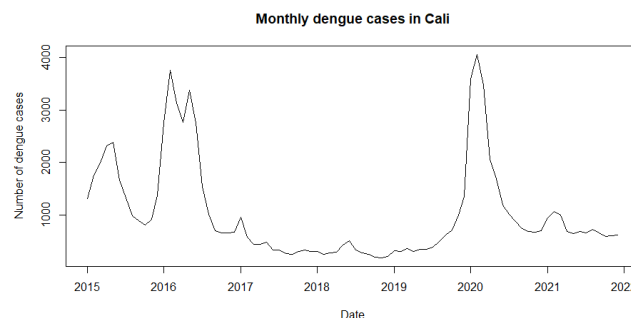
205 It should be noted that information from 2005 up to 2022 was used in order to have more
206 data for the regression model.

207 **3.2. Lagged correlation analysis and election of variables.**

208 Fig 3 shows the time series of dengue incidence in Cali. It can be observed that dengue
209 incidence is characterized by long periods of relative tranquility, where there are a few
210 dengue cases, and some peaks where incidence shoots up and reaches much higher values.
211 Below the relationship between this behavior and the aforementioned climatic variables.

212

Fig. 3 Monthly dengue cases in Cali

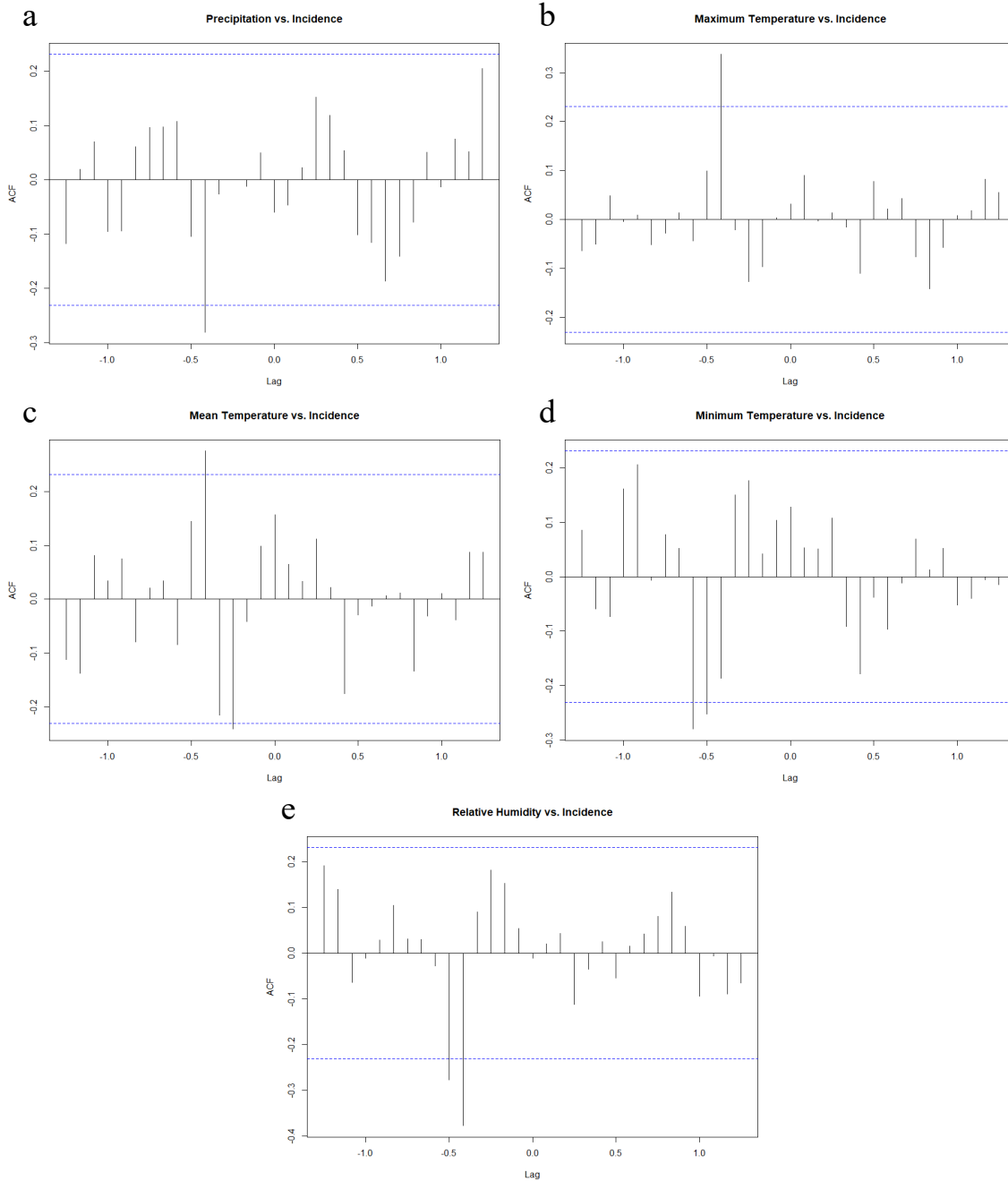


213

214 Using R, the correlation of each variable with dengue incidence was studied. However, the
215 nature of the vector implies the need of considering a lag. In fact, later on we verify that the

216 lagged variables show higher correlation with dengue incidence than without any lag. Fig 4
217 shows the lagged cross correlation analysis for all the climatic variables.

218 **Fig. 4 Lagged correlation analysis**



219 For precipitation (Fig 4. a) there's significant negative correlation at -0.4 lag, meaning
220 between incidence and the precipitation of 5 months in the past. For maximum temperature

221 (Fig. 4 b) the most significant correlation is around a -0.4 lag, resulting in a 4-month lag. The
 222 relation is direct. For average temperature (Fig. 4 c) there are two lags with significant
 223 correlation, a direct correlation with a 5-month lag and an inverse correlation with a 3-month
 224 lag. For minimum temperature the significant correlation is negative at a 7-month lag. And
 225 for relative humidity there is negative significant correlation at a 5-month lag. To summarize,
 226 the chosen variables after the lagged correlation analysis are Precipitation(t-5), Maximum
 227 Temperature(t-5), Mean Temperature(t-3), Mean Temperature(t-5), Minimum Temperature(t-
 228 7) and Relative Humidity(t-5). It must be noted that before doing this analysis the first
 229 difference of ever variables was calculated in order to guarantee stationarity and avoid
 230 spurious correlation.

231 3.3. Principal component analysis (ACP)

232 The multicollinearity problem must be addressed since, naturally, climatic variables are
 233 correlated amongst themselves. Table 1 shows de correlation matrix of the climatic variables.

234 **Table 1 Correlation matrix of the climatic variables**

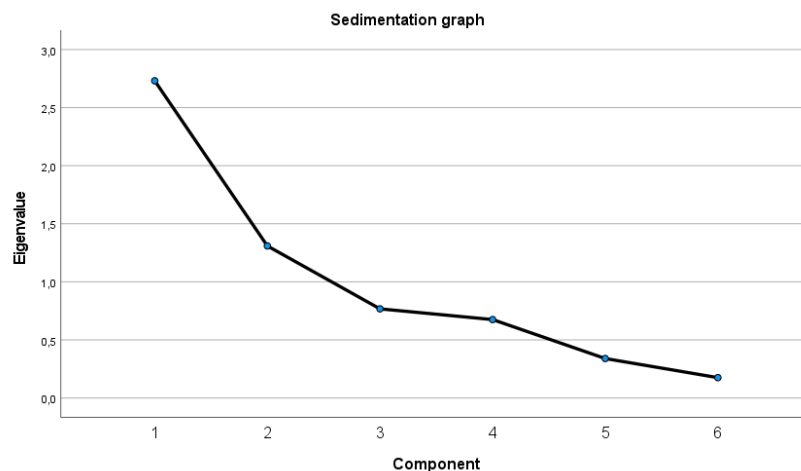
Correlation matrix							
		Precipitation (t-5)	Maximum Temperature (t-5)	Mean Temperature (t-5)	Mean Temperature (t-3)	Minimum Temperature (t-7)	Relative Humidity (t-5)
Correlation	Precipitation (t-5)	1,000	-,388	-,387	,012	,116	,555
	Maximum Temperature (t-5)	-,388	1,000	,766	,065	-,054	-,605
	Mean Temperature (t-5)	-,387	,766	1,000	,119	,096	-,698

	Mean Temperature (t-3)	,012	,065	,119	1,000	,238	,060
	Minimum Temperature (t-7)	,116	-,054	,096	,238	1,000	,174
	Relative humidity (t-5)	,555	-,605	-,698	,060	,174	1,000
Sig. (unilateral)	Precipitation (t-5)		<.001	<.001	,460	,167	<.001
	Maximum Temperature (t-5)	,000		,000	,295	,327	,000
	Mean Temperature (t-5)	,000	,000		,161	,212	,000
	Mean Temperature (t-3)	,460	,295	,161		,023	,309
	Minimum Temperature (t-7)	,167	,327	,212	,023		,073
	Relative humidity (t-5)	,000	,000	,000	,309	,073	

235 Fig 5 shows the associated variance to each of the principal components. According to this,
 236 we pick the first to components which explain 67,3% of the variance.

237

Fig. 5 Sedimentation graph of the ACP

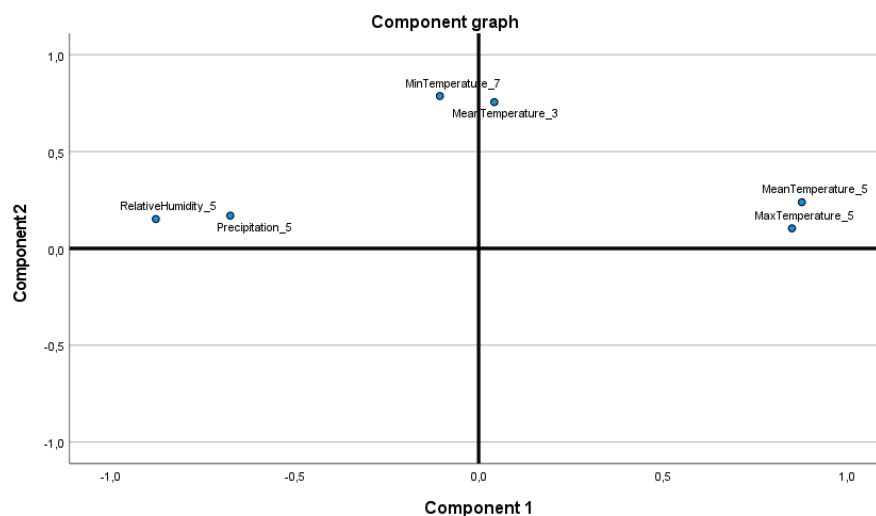


238

239 Fig 6 reveals that the variables can be separated in two groups. Therefore, instead of using
240 the principal components as predictor variables, we choose one variable for each group
241 making the regression easier to understand and conclude. The selected variables are Mean
242 Temperature(t-5) and Mean Temperature(t-3).

243

Fig. 6 Principal components in rotated space



244

245 3.4. Regression models

246 Table 2 presents a comparison of the three regression models used in this work including
 247 deviance, Akaike information criterion (AIC) and the p-values of the adjusted models for
 248 every variable. For every model the best variable choice was both mean temperature lag
 249 variables and their interaction.

250 **Table 2. Comparison of the regression models.**

Model	Residual deviance	AIC	P-VALUES			
			Intercept	Mean Temperature (t-5)	Mean temperature (t-3)	Interaction
Poisson	34224.0	34826.0	2.00E-16	2.00E-16	2.00E-16	2.00E-16
Negative binomial	76.388	1095.80	0.00898	0.00893	0.01	0.01249
PIG	1075.814	1085.81	0.0439	0.0423	0.0463	0.0539

251 Clearly, the Poisson model is not good, probably because dengue incidences is over
 252 dispersed, its mean is 981.6 and its variance is 950046.6. Therefore, this model was
 253 discarded. Regarding the negative binomial and the PIG models, the negative binomial is
 254 chosen because its deviance is better and shows a slightly better adjustment to the data as
 255 seen in the p-values. As Table 3 shows, all the variables and the intercept were significant for
 256 the negative binomial model.

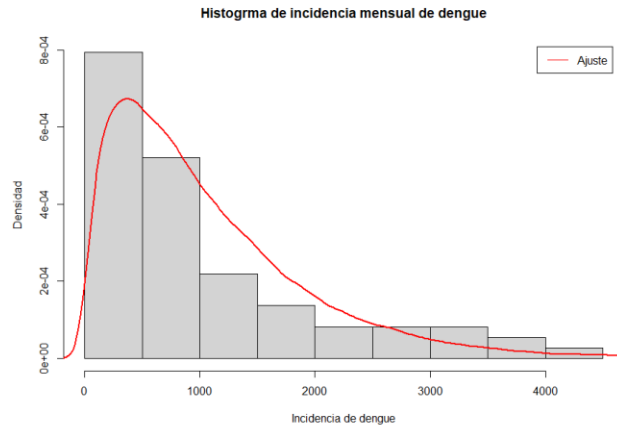
257 **Table 3 Coefficients of the negative binomial model**

Coefficients	Estimation	Std. Error	z value	Pr(> z)
(Intercept)	-225.2824	86.2199	-2.613	0.00898
Mean Temperature(t-5)	8.9056	3.4062	2.615	0.00893
Mean Temperature(t-3)	8.8248	3.4262	2.576	0.01

Interaction	-0.338	0.1353	-2.498	0.01249
-------------	--------	--------	--------	---------

258 We can also verify that the data resemble a negative binomial distribution as Fig 7 shows.

259 **Fig. 7 Frequency distribution**



260

261 The resulting model can be expressed as follows:

262
$$Dengue = e^{(-225.2824 + 8.9056 MeanTemp_5 + 8.8248 MeanTemp_3 - 0.338 MeanTemp_5 MeanTemp_3)}$$

263 Fig 8 displays both real data and data predicted by the negative binomial regression model.

264 The fit is good since the predicted values (red) resemble the real data (black) in the time

265 series as well as the histogram (Fig 9). However, it's clear that the fit could be better, for

266 example, the peaks of 2016 and 202 aren't predicted correctly, while the dengue cases on the

267 low incidence seasons are overestimated. It must be kept in mind that 2020 was an atypical

268 year due to the Covid-19 pandemic.

Fig. 8 Adjusted regression model for dengue incidence

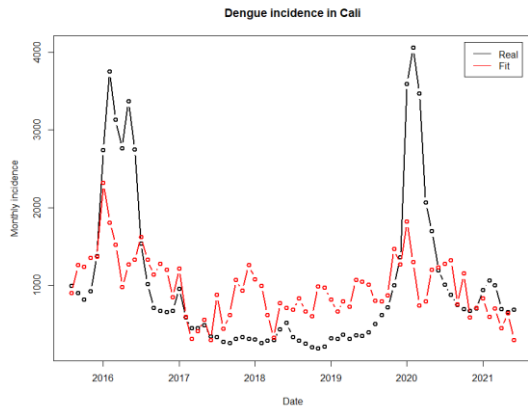
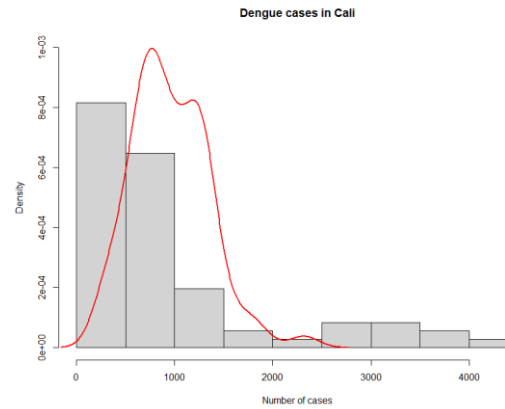


Fig. 9 Histogram of dengue incidence



269

270 Knowing that the average temperature for the data set is 25°C, the coefficients can be used
271 for the following interpretations: If the average monthly temperature in Cali increases by one
272 degree, a 45% increase in dengue cases is expected after 3 months. Additionally, after 5
273 months, the same increase of one degree will imply a 64% increase in dengue incidence. This
274 is particularly worrying considering the temperature rise tendencies because of global
275 warming and the recent arrival of El Niño phenomenon.

276 4. Discussion

277 The results obtained on the data imputation process shows that this methodology is a feasible
278 and efficient way of estimating missing data.

279 The negative binomial regression model was found to be the best for predicting monthly
280 dengue incidence in Cali using mean temperature with 3 and 5 months lags. This implies that
281 the effect of a change in temperature takes about 3 months to be visible in dengue incidence.

282 This could be attributed to the biological processes of *Aedes aegypti* such as maturation and

283 reproduction which are known to be affected by the climatic variables. Taking into
284 consideration that the life span of a female mosquito is about 6 weeks, we can conclude that
285 it takes at least two generation of mosquitoes for the dengue incidence to be affected by a
286 change in temperature.

287 In other matters, the absence of other climatic variables from the regression model does not
288 necessarily imply that they have no effect over dengue incidence, seeing that the naturally
289 have a big correlation with temperature. For instance, both precipitation and relative humidity
290 are inversely proportional to temperature, as the ACP shows.

291 Forecasts are worrisome, because an increase in average temperature could transform into a
292 significant increase in dengue cases having serious repercussions on public health.

293 Finally, its recommended to deepen on this type of predictive studies. This research has some
294 limitations that can be improved upon on further work. Firstly, there isn't a lot of climatic
295 data. Only one meteorological station with enough data was found, and even then, this data
296 was incomplete and hence an imputation processes was necessary. Secondly, expanding the
297 time period of the study is recommended in order to obtain more robust results, but this will
298 imply recollecting more data for both dengue and the climatic variables. And finally, this
299 work didn't take into consideration the autoregressive component of dengue (present dengue
300 incidence depends on the incidence of past months). This could lead to more accurate
301 predictions.

302 **5. References**

303 1. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global
304 distribution and burden of dengue. *Nature*. 2013 Apr 25;496(7446):504–7.

- 305 2. Organización Mundial de la Salud. DENGUE guidelines for diagnosis, treatment,
306 prevention and control [Internet]. 2009. Available from: www.who.int/tdr
- 307 3. World Mosquito Program. Conoce las últimas novedades de World Mosquito Program
308 en Cali. [Internet]. World Mosquito Program. 2023 [cited 2023 Mar 14]. Available
309 from: [https://www.worldmosquitoprogram.org/es/avances-nivel-](https://www.worldmosquitoprogram.org/es/avances-nivel-mundial/colombia/cali)
310 [mundial/colombia/cali](https://www.worldmosquitoprogram.org/es/avances-nivel-mundial/colombia/cali)
- 311 4. Márquez Benitez Y, Monroy Cortes KJ, Martínez Montenegro EG, Peña García VH,
312 Monroy Díaz ÁL. Influencia de la temperatura ambiental en el mosquito *Aedes* spp y
313 la transmisión del virus del dengue. *CES Med.* 2019;33(1):42–50.
- 314 5. Nelson MJ. *Aedes Aegypti: Biología y Ecología*. Washington D.C.: Organización
315 Panamericana de la Salud; 1986. 4–34 p.
- 316 6. Britos Molinas MB. Implicancia del cambio climático en la aparición del riesgo de
317 transmisión de dengue por *Aedes aegypti* y su adaptación a las condiciones
318 ambientales locales. *Rev la Soc Científica del Paraguay.* 2022 Jun 16;27(1):120–7.
- 319 7. Morin CW, Comrie AC, Ernst K. Climate and dengue transmission: Evidence and
320 implications. *Environ Health Perspect.* 2013;121(11–12):1264–72.
- 321 8. Rúa-Uribe GL, Suárez-Acosta C, Chauca J, Ventosilla P, Almanza R. Modelado del
322 efecto de la variabilidad climática local sobre la transmisión de dengue en Medellín
323 (Colombia) mediante análisis de series temporales. *Biomedica.* 2013
324 Sep;33(SUPPL.1):142–52.
- 325 9. Ordóñez-Sierra R, Gómez-Albores MÁ, Mastachi-Loza CA, Díaz-Delgado C,
326 Manzano-Solís R, Medina-Torres I. Análisis de cambio en las condiciones climáticas

- 327 y su relación con casos de dengue en la parte centro de México. Digit
328 Ciencia@UAQRO [Internet]. 2019 [cited 2023 Nov 8];12(1):96–105. Available from:
329 <https://revistas.uaq.mx/index.php/ciencia/article/view/20>
- 330 10. Rubio-Palis Yasmin, Perez-Ybarra LM, Infante-Ruiz M, Comach G, Urdaneta-
331 Márquez L. Influencia de las variables climáticas en la casuística de dengue y la
332 abundancia de *Aedes aegypti* (Diptera: Culicidae) en Maracay, Venezuela. Boletín
333 Malariol y Salud Ambient [Internet]. 2011 Nov;51:145–58. Available from:
334 [http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1690-](http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1690-46482011000200004&nrm=iso)
335 [46482011000200004&nrm=iso](http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1690-46482011000200004&nrm=iso)
- 336 11. Jeelani S, Sabesan S. *Aedes* vector population dynamics and occurrence of dengue
337 fever in relation to climate variables in Puducherry, South India. *Int J Curr Microbiol*
338 *Appl Sci*. 2013;2(12):313–22.
- 339 12. Liyanage P, Tissera H, Sewe M, Quam M, Amarasinghe A, Palihawadana P, et al. A
340 spatial hierarchical analysis of the temporal influences of the el niño-southern
341 oscillation and weather on dengue in Kalutara District, Sri Lanka. *Int J Environ Res*
342 *Public Health*. 2016 Nov 1;13(11).
- 343 13. Rúa Uribe GL, Calle Londoño DA, Rojo Ospina RA, Henao Correa EA, Sanabria
344 González WH, Suárez Acosta CDR. Influencia del evento climático El Niño sobre la
345 dinámica de transmisión de dengue en Medellín, Antioquia, Colombia. *IATREIA*.
346 2012;25(4):314–22.
- 347 14. Muñoz E, Poveda G, Arbeláez MP, Vélez ID. Spatiotemporal dynamics of dengue in
348 Colombia in relation to the combined effects of local climate and ENSO. *Acta Trop*.

- 349 2021 Dec 1;224.
- 350 15. Hossain S. Generalized Linear Regression Model to Determine the Threshold Effects
351 of Climate Variables on Dengue Fever: A Case Study on Bangladesh. *Can J Infect Dis*
352 *Med Microbiol.* 2023;2023.
- 353 16. Ferreira Saraiva E, Piedade Vigas V, Flesch M, Gannon M, de Bragança Pereira CA.
354 Modeling Overdispersed Dengue Data via Poisson Inverse Gaussian Regression
355 Model: A Case Study in the City of Campo Grande, MS, Brazil. *Entropy.* 2022 Sep
356 1;24(9).
- 357 17. Pinto E, Coelho M, Oliver L, Massad E. The influence of climate variables on dengue
358 in Singapore. *Int J Environ Health Res.* 2011 Dec;21(6):415–26.
- 359 18. Kakarla SG, Caminade C, Mutheneni SR, Morse AP, Upadhyayula SM, Kadiri MR, et
360 al. Lag effect of climatic variables on dengue burden in India. *Epidemiol Infect.*
361 2019;147.
- 362 19. de Almeida Pereira MM, Ângelo Barrêto FF, Rufino de Freitas J, Mendes de
363 Vasconcelos J, Alves da Silva AS, Rocha Moreira G, et al. Dengue in Brazil:
364 Regression and probabilistic modeling. *Int J Dev Res.* 2022 Jun 28;12(06):56938–43.
- 365 20. Fathurahman M, Purnamasari I, Prangga S. Negative binomial regression analysis on
366 dengue hemorrhagic fever cases in East Kalimantan Province. *AIP Conf Proc*
367 [Internet]. 2022 Oct 11;2668(1):070002. Available from:
368 <https://doi.org/10.1063/5.0111999>
- 369 21. Abhinandithe SK, Madhu B, Balasumbaramanian S. Negative binomial regression
370 modeling to assess the influence of climatic factors on the dengue incidence during an

- 371 epidemic in Mysore district. *Int J Mosq Res* [Internet]. 2021;8(2):19–29. Available
372 from: <http://www.dipterajournal.com>
- 373 22. Sinaga JP, Sinulingga U. Poisson Regression Modeling Case Study Dengue Fever in
374 Medan City in 2019. *J Math Technol Educ* [Internet]. 2021;1(1):94–102. Available
375 from: <https://talenta.usu.ac.id/jomte>
- 376 23. Alcaldía de Santiago de Cali. En 2022, Cali redujo los casos de infección y muerte por
377 dengue [Internet]. Alcaldía de Santiago de Cali. 2022 [cited 2023 Mar 14]. Available
378 from: <https://www.cali.gov.co/salud/publicaciones/173546/en-2022-cali-redujo-los-casos-de-infeccion-y-muerte-por-dengue/#:~:text=El resultado de 2022 con,se han reportado 2741 casos>
379
380
- 381 24. Alcaldía de Santiago de Cali. Geografía de Cali [Internet]. 2014 [cited 2023 Nov 21].
382 Available from:
383 https://www.cali.gov.co/informatica/publicaciones/106104/geografia_de_cali/
- 384 25. Jin Z, Shang J, Zhu Q, Ling C, Xie W, Qiang B. RFRSF: Employee Turnover
385 Prediction Based on Random Forests and Survival Analysis. *Lect Notes Comput Sci*
386 (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2020;12343
387 LNCS:503–15.
- 388 26. Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: A
389 large-scale benchmark experiment. *BMC Bioinformatics*. 2018;19(1):1–14.
- 390 27. Fairos W, Azaki W, Alias L, Wah Y. Modelling Dengue Fever (DF) and Dengue
391 Haemorrhagic Fever (DHF) Outbreak Using Poisson and Negative Binomial Model.
392 *World Acad Sci Eng Technol*. 2010;62:903–8.

- 393 28. Hilbe JM. Modeling Count Data [Internet]. Cambridge University Press; 2014.
394 Available from:
395 <https://www.cambridge.org/core/product/identifier/9781139236065/type/book>
- 396 29. Arriagada P, Karelavic B, Link O. Automatic gap-filling of daily streamflow time
397 series in data-scarce regions using a machine learning algorithm. J Hydrol. 2021 Jul
398 1;598.

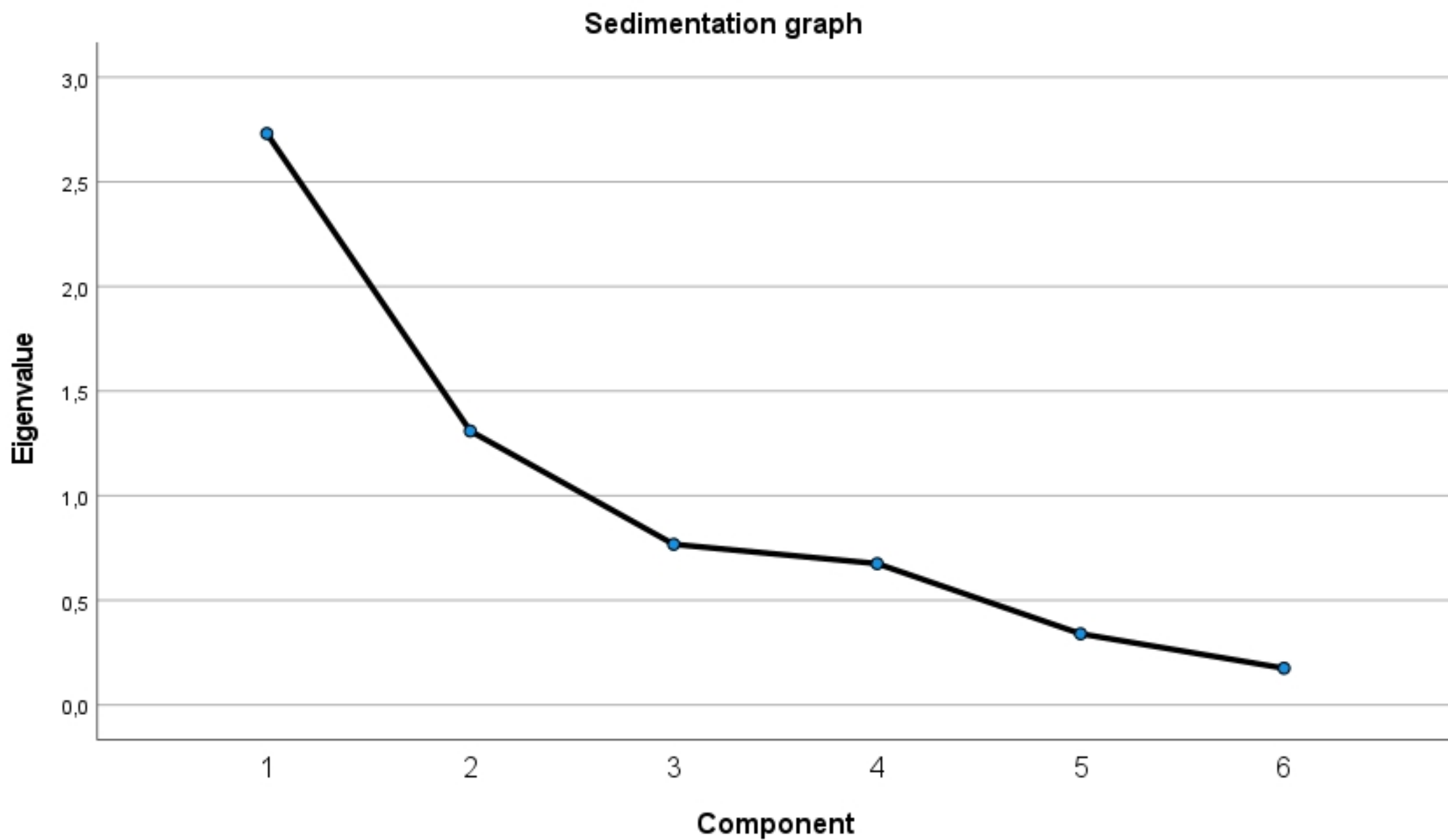


Fig 5

Dengue incidence in Cali

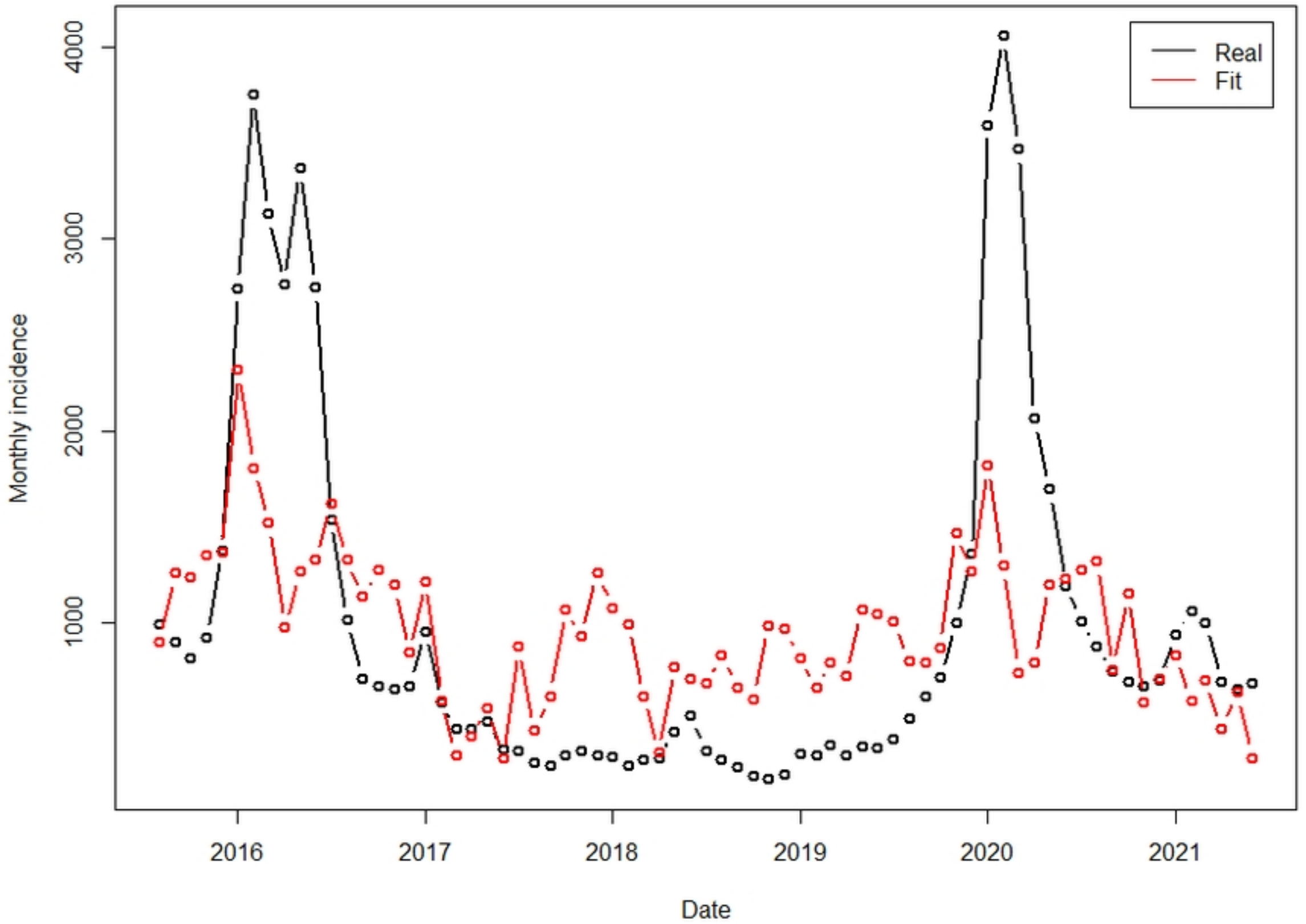


Fig 8

Dengue cases in Cali

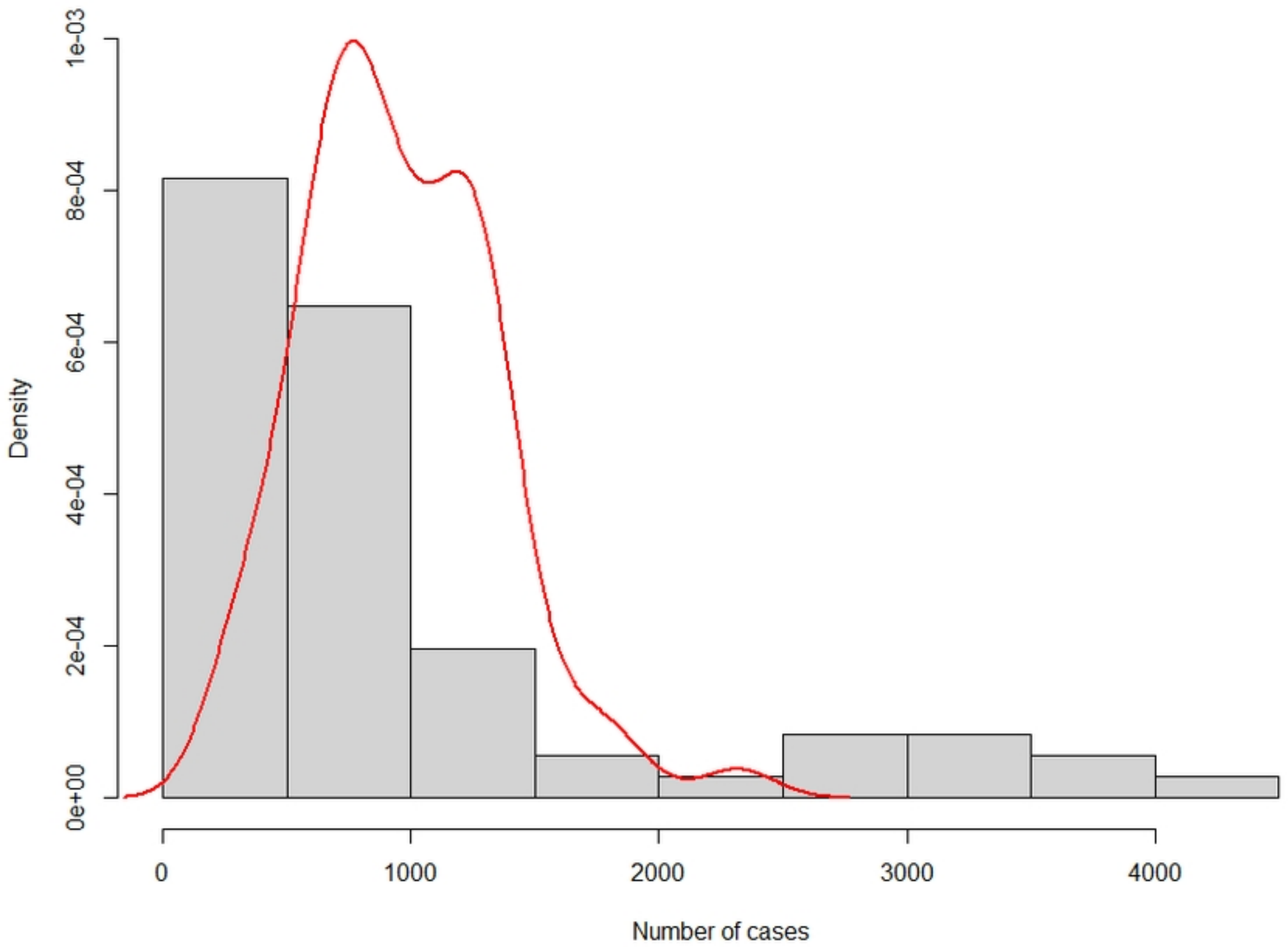


Fig 9

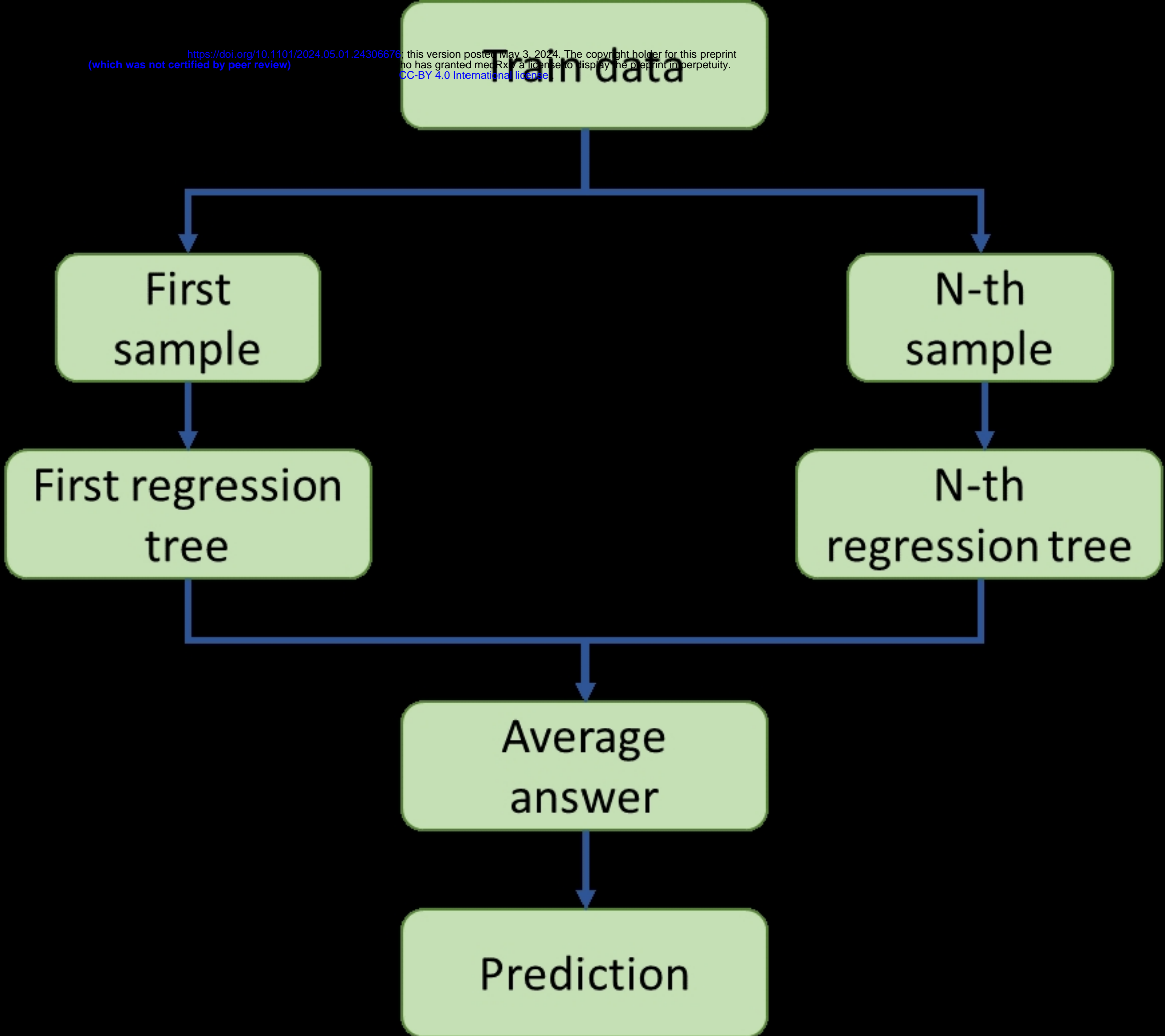


Fig 1

Monthly dengue cases in Cali

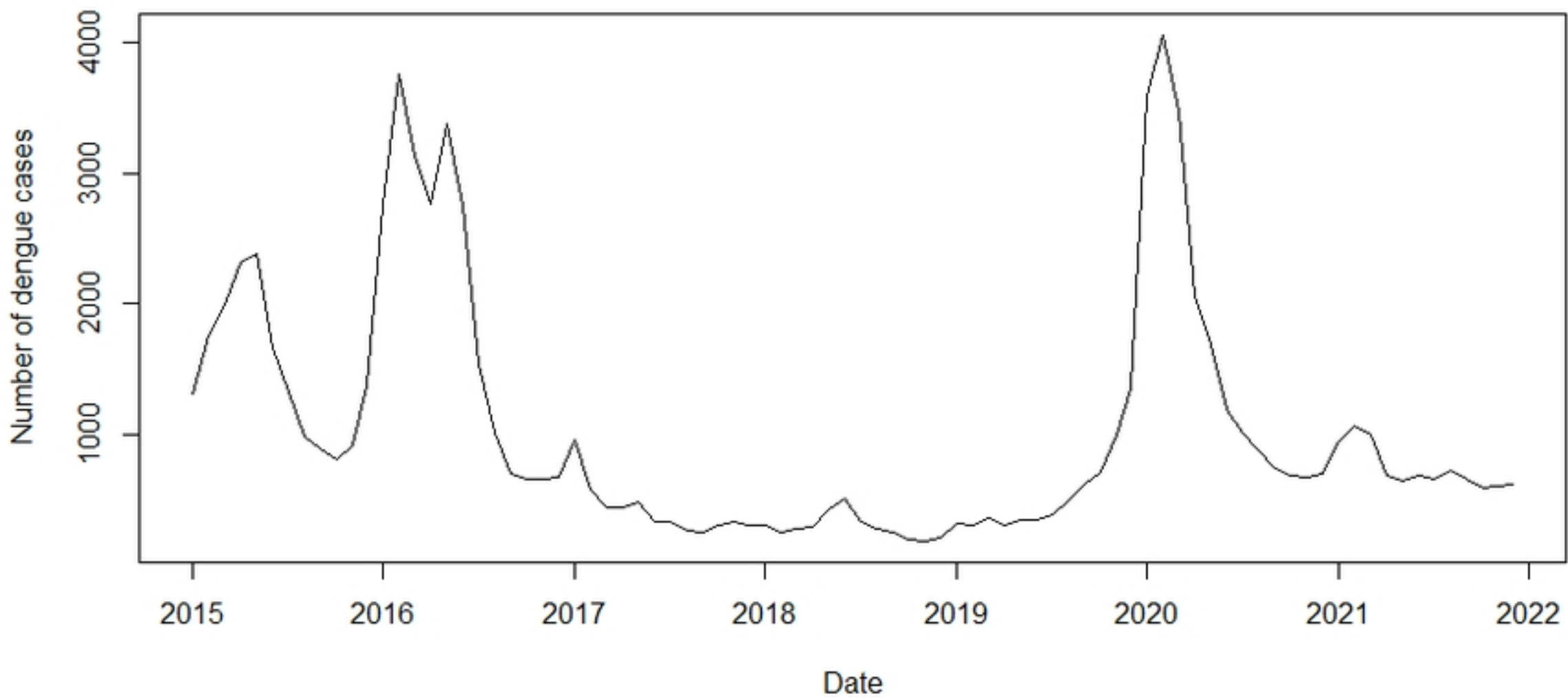
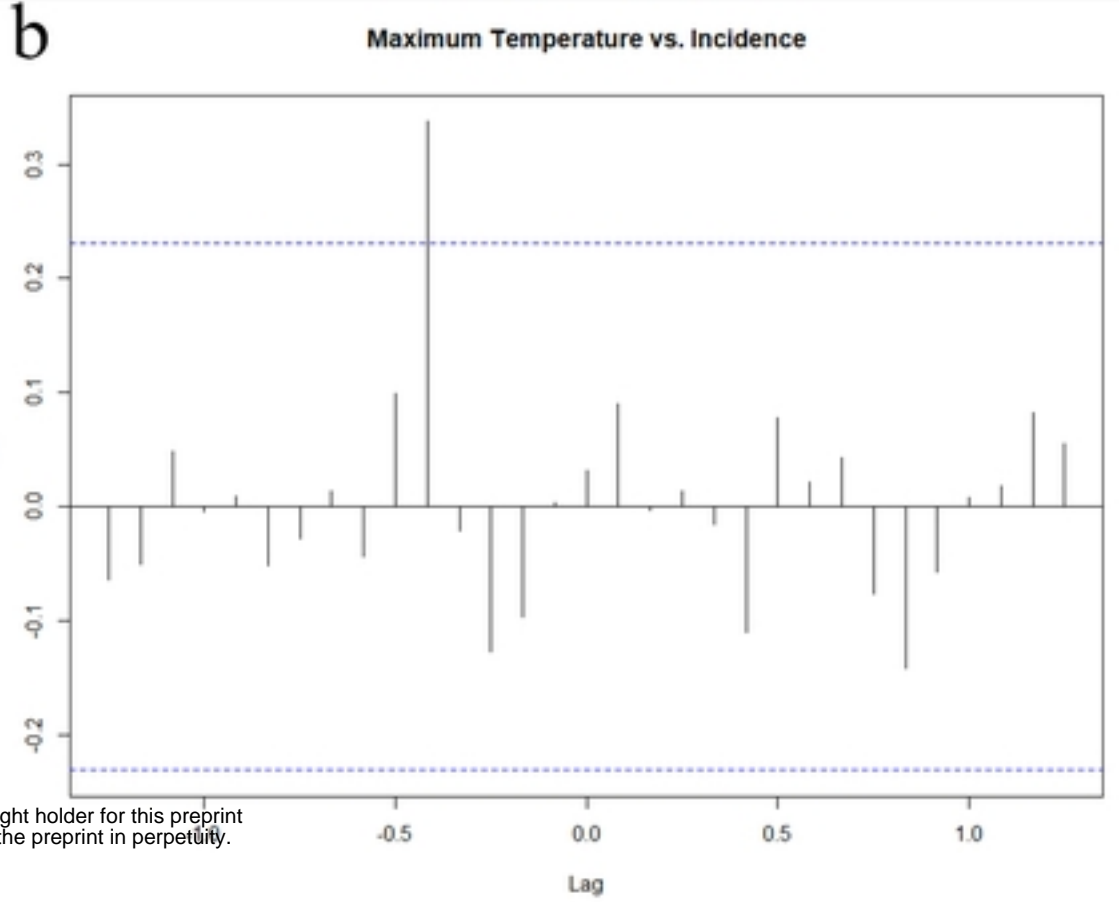
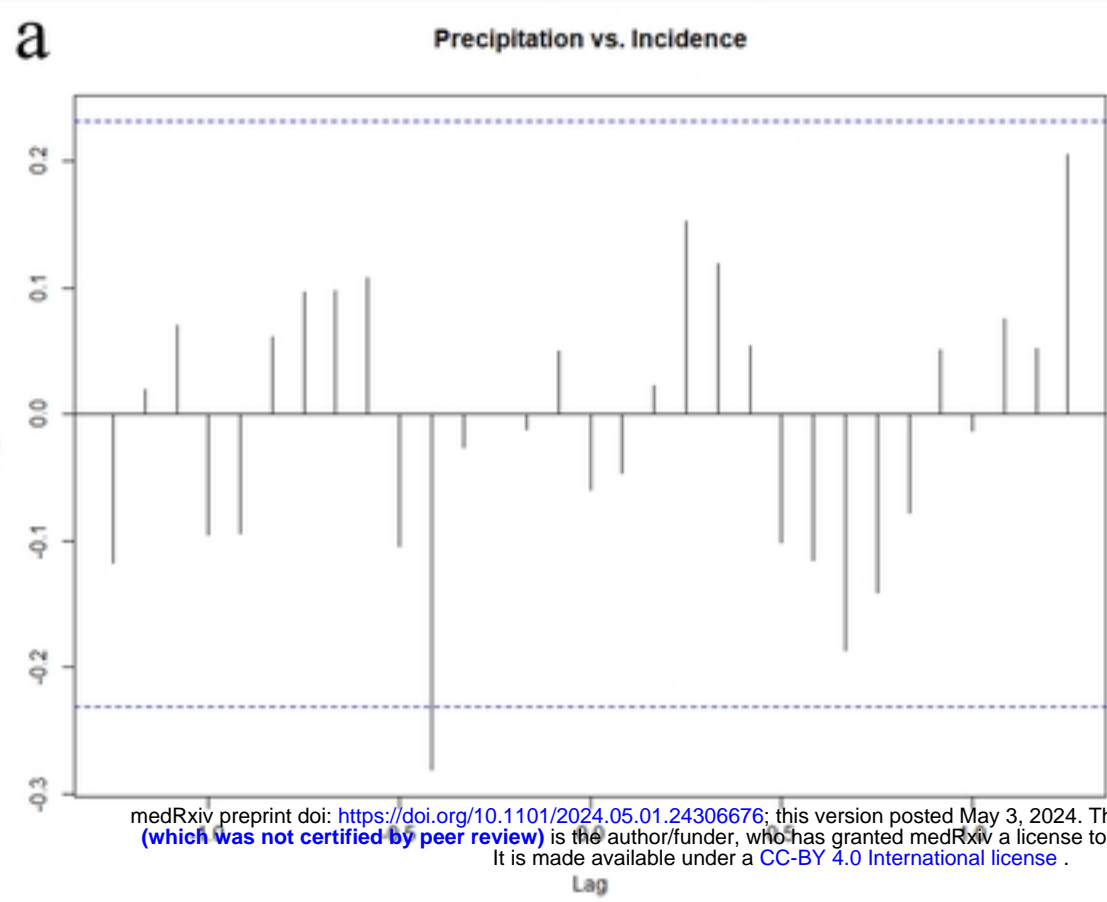


Fig 3



medRxiv preprint doi: <https://doi.org/10.1101/2024.05.01.24306676>; this version posted May 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

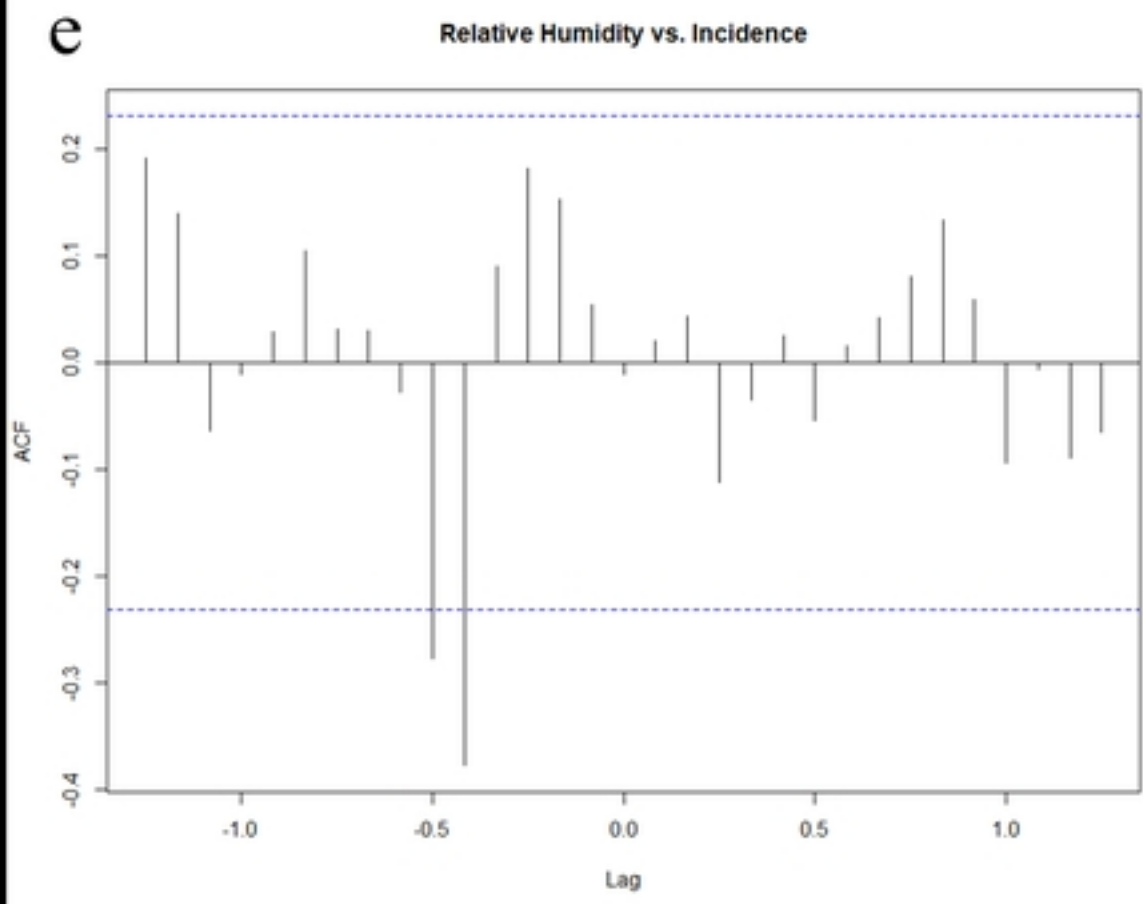
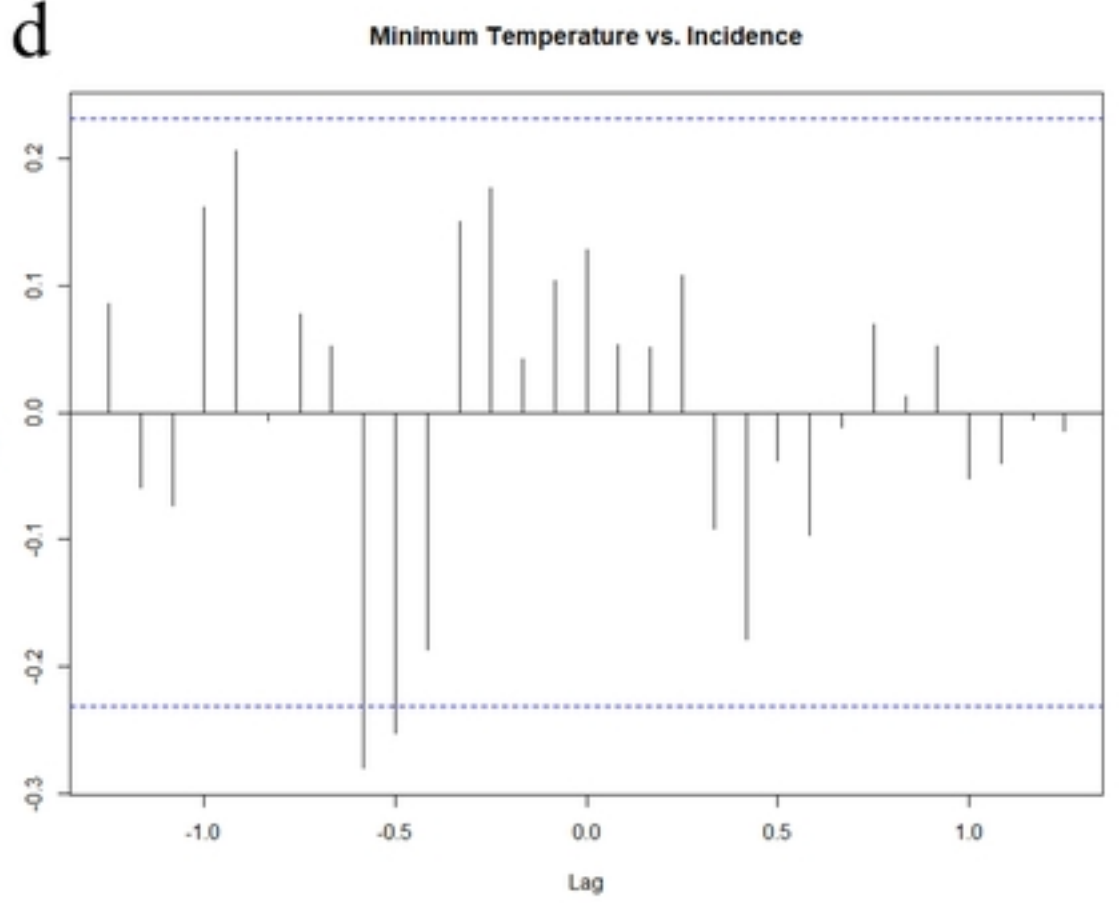
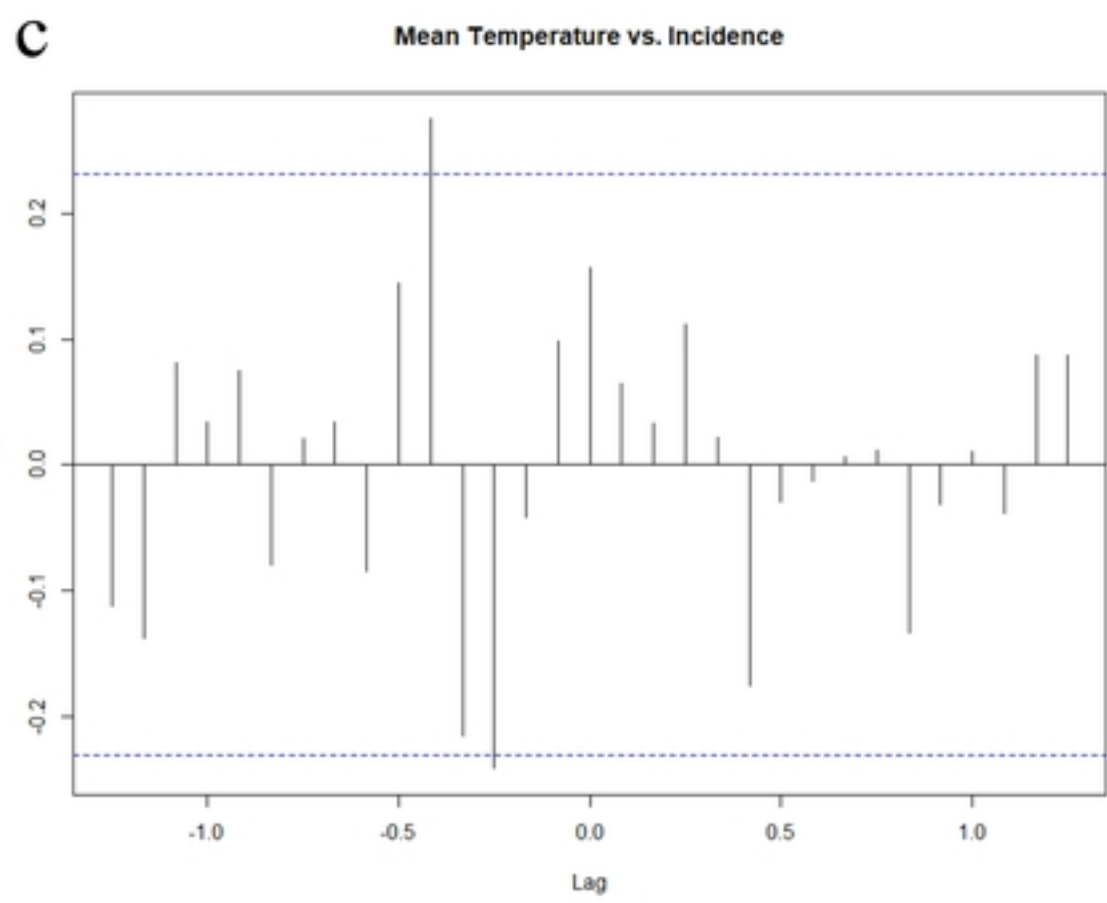
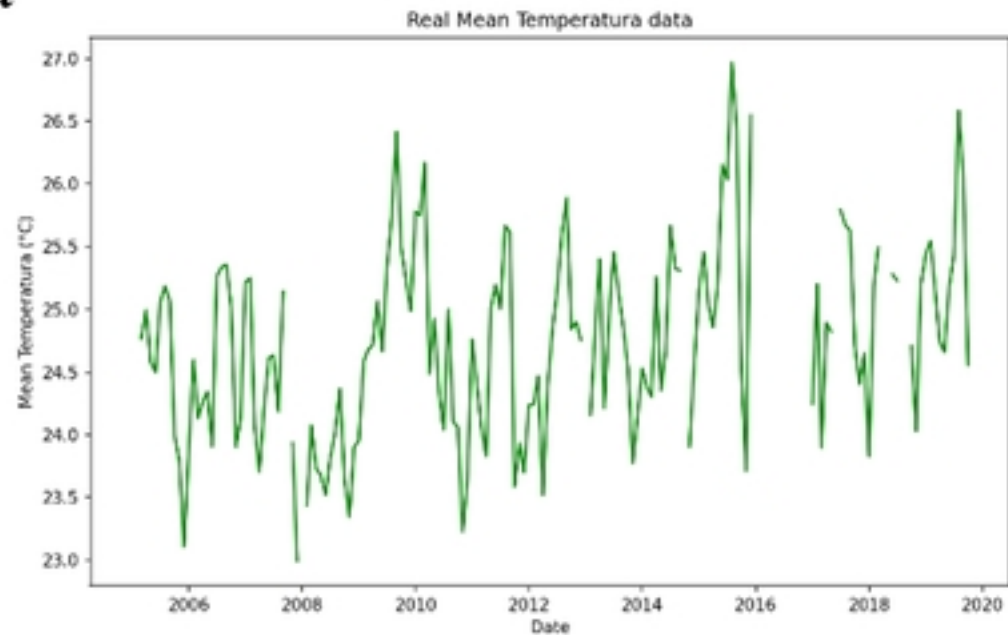
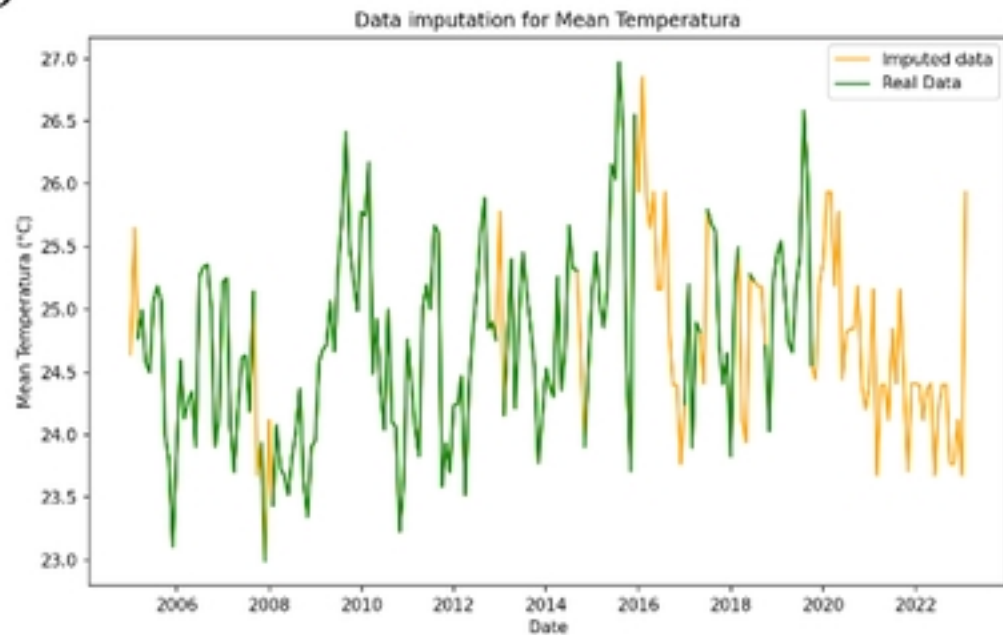


Fig 4

a**b****Fig 2**