








## Article

# Multiplicative Mixed-Effects Modelling of Dengue Incidence: An Analysis of the 2019 Outbreak in the Dominican Republic

Adelaide Freitas <sup>1,2,†</sup> , Helena Sofia Rodrigues <sup>2,3,\*,†</sup> , Natália Martins <sup>1,2,\*,†</sup> , Adela Iutis <sup>1</sup> ,  
Michael A. Robert <sup>4,5,†</sup> , Demian Herrera <sup>6</sup>  and Manuel Colomé-Hidalgo <sup>7</sup> 

- <sup>1</sup> Department of Mathematics, University of Aveiro, 3810-193 Aveiro, Portugal  
<sup>2</sup> Center for Research and Development in Mathematics and Applications (CIDMA), University of Aveiro, 3810-193 Aveiro, Portugal  
<sup>3</sup> Business School, Instituto Politécnico de Viana do Castelo, 4930-600 Valença, Portugal  
<sup>4</sup> Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, USA  
<sup>5</sup> Center for Emerging, Zoonotic, and Arthropod-Borne Pathogens, Virginia Tech, Blacksburg, VA 24060, USA  
<sup>6</sup> Centro de Investigación en Salud Dr. Hugo Mendoza, Hospital Pediátrico Dr. Hugo Mendoza, Santo Domingo 10117, Dominican Republic  
<sup>7</sup> Instituto Tecnológico de Santo Domingo (INTEC), Santo Domingo 10602, Dominican Republic  
\* Correspondence: sofia.rodrigues@esce.ipv.pt (H.S.R.); natalia@ua.pt (N.M.)  
† These authors contributed equally to this work.

**Abstract:** Dengue is a vector-borne disease that is endemic to several countries, including the Dominican Republic, which has experienced dengue outbreaks for over four decades. With outbreaks growing in incidence in recent years, it is becoming increasingly important to develop better tools to understand drivers of dengue transmission. Such tools are critical for providing timely information to assist healthcare authorities in preparing human, material, and medical resources for outbreaks. Here, we investigate associations between meteorological variables and dengue transmission in the Dominican Republic in 2019, the year in which the country's largest outbreak to date occurred. We apply generalized linear mixed modelling with gamma family and log link to model the weekly dengue incidence rate. Because correlations in lags between climate variables and dengue cases exhibited different behaviour among provinces, a backward-type selection method was executed to find a final model with lags in the explanatory variables. We find that in the best models, meteorological conditions such as temperature and rainfall have an impact with a delay of 2–5 weeks in the development of an outbreak, ensuring breeding conditions for mosquitoes.

**Keywords:** dengue; Dominican Republic; climate variables; lags; generalized linear mixed models

**MSC:** 92B15; 62P10; 62J12



**Citation:** Freitas, A.; Rodrigues, H.S.; Martins, N.; Iutis, A.; Robert, M.A.; Herrera, D.; Colomé-Hidalgo, M. Multiplicative Mixed-Effects Modelling of Dengue Incidence: An Analysis of the 2019 Outbreak in the Dominican Republic. *Axioms* **2023**, *12*, 150. <https://doi.org/10.3390/axioms12020150>

Academic Editors: Pasquale Palumbo, Nuno Bastos, Touria Karite and Amir Khan

Received: 19 December 2022

Revised: 10 January 2023

Accepted: 21 January 2023

Published: 1 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Dengue is one of the most significant mosquito-borne diseases to threaten human populations, particularly in tropical and subtropical regions. The number of dengue cases reported to the World Health Organization (WHO) has increased sharply from less than 0.5 million in 2000 to 5.2 million in 2019, and the number of dengue-induced deaths increased from 960 in 2000 to 4032 in 2015 [1], leading WHO to name dengue as one of the ten biggest threats to global health in 2019 [2]. Given the global increases in dengue in recent years, it is increasingly important to develop tools to better understand drivers of dengue transmission and to predict future outbreaks.

One country where dengue has long been endemic is the Dominican Republic. In the past decade, however, dengue outbreaks have grown in incidence, with the 2019 outbreak being the largest outbreak in the country to date [3,4]. In fact, 2019 was the year in which the WHO recorded the highest number of global dengue cases ever to occur within a

year [1], suggesting that dengue in the Dominican Republic is mirroring global trends. In an exploratory analysis of dengue in the Dominican Republic, Iutis et al. [3] showed there is no single meteorological, demographic or geographic factor that affects the incidence rate of dengue. They instead suggest that a combination of different factors could be responsible for increases in dengue cases. Among these factors is climate, which plays a very important role in dengue transmission and in the life cycle of the mosquitos that transmit dengue virus. For example, it is well known that the mosquito must have certain meteorological conditions to survive and reproduce [5,6]. Herein, we aim to characterize relationships between dengue cases and meteorological variables such as temperature, humidity, and precipitation by considering the 2019 outbreak of dengue in multiple provinces of the Dominican Republic. It is important to study not only the impact of climate variables on dengue transmission, but also lags between dengue cases and these variables because there are inherent lags in the dengue transmission process that arise from the mosquito life cycle. To that end, we also explore the relevance of time between meteorological conditions and reported dengue incidence by studying lags between climate and dengue variables.

By using 2019 dengue case data collected by hospitals in geographically distinct areas of the Dominican Republic, we implement gamma generalized linear mixed models (gamma-GLMM) to model relationships between dengue incidence rate and climatic variables, such as temperature, humidity, and precipitation. We emphasize here that 2019 is an important year in the evolution of dengue in the Dominican Republic and globally because the highest number of dengue cases ever reported both in the country and globally was in 2019 [1], and an investigation such as the present one will contribute to a better understanding of the drivers of this large outbreak.

This paper is organized as follows. First, we review literature of recent research on dengue. In Section 3, we introduce the response variable and discuss possible explanatory variables. We analyze the effect of lags between variables by studying correlations between the response variable and the meteorological variables with delays. Section 4 describes the gamma-GLMM method implemented for this study. In Section 5, we present two regression models and their results and discuss implications for modelling the weekly dengue incidence rate. Finally, in Section 6, we present some conclusions and directions for future work.

## 2. Literature Overview

### 2.1. Disease Transmission

Dengue virus is transmitted to humans by female mosquitoes, mainly of the species *Aedes aegypti* and *Aedes albopictus*. There are four strains of the dengue virus, and people can be infected with the virus more than once [7]. Infection in humans begins when an infectious female mosquito bites a human and injects dengue virus from one strain into the blood of the human host. Then the dengue virus develops and causes symptomatic or asymptomatic infection in humans. Symptoms of the disease can range from mild forms such as sudden fever, severe headache, nausea, vomiting, myalgia, and skin erythema, to more severe forms including dengue hemorrhagic fever and dengue shock syndrome. Severe dengue can cause death due to plasma leakage, fluid accumulation, severe bleeding, and respiratory failure [1].

On average, dengue infection persists for approximately 2 weeks [8]. The infected person has permanent immunity to the strain of dengue virus that caused the illness and temporary immunity to the other three strains. It should be noted that, in many cases, a second infection with a different strain of dengue virus can lead to a more virulent form of the disease [9]. Dengue virus transmission depends on four factors: the presence of the virus, the human host, the mosquito vector, and the suitability of environmental conditions [10]. With regard to environmental conditions, the transmission of the dengue virus is influenced by several factors, such as temperature, precipitation, relative humidity, and rapid urbanization [1].

## 2.2. Disease Control

To date, there are no effective antiviral therapies and the only treatment is to control the symptoms with medication. Vaccination is a measure of reduced effectiveness because currently there is only one licensed dengue vaccine which has several limitations. In particular, it can only be administered to people between 9 and 45 years old who have already been infected with one of the dengue viruses [1,11]. Vector control is the only available strategy against dengue. To this end, it is possible to implement measures including the use of insecticides and educational campaigns. Although insecticides have been effective in controlling dengue, increasing trends in mosquito-borne diseases may indicate an increase in insecticide resistance or ineffectiveness in controlling dengue transmission. Therefore, it is of great importance to understand mechanisms of resistance and the susceptibility of the mosquitoes to insecticides in order to develop more effective *Aedes* mosquito-control methods [12]. Educational campaigns are of great importance in preventing and controlling the spread of dengue. It is very important that the population recognize the symptoms of dengue, to be aware of the importance of having medical treatment in case of severe dengue, and to know how to control populations of the *Aedes* mosquito. In [13] the authors concluded that the population of Sri Lanka in 2019 has better awareness of dengue prevention compared to awareness of dengue mortality and dengue management. This study on knowledge, attitudes, and practices regarding dengue fever identified as a weak point the patient awareness of the patient's role in the management of dengue and identification of warning signs that precede hospitalization. If dengue hemorrhagic fever is detected early, the mortality is 2–5% but it can increase to 20% if there is no immediate treatment.

## 2.3. Dengue Modelling

Simulation models are useful for understanding the drivers and spread of dengue and for helping to understand the efficacy of potential control methods. Many simulation model studies use dynamical models based on ordinary differential equations [14–16]. However, in general, these models do not describe the effects that arise from delays between drivers and reported cases. There are inherent delays in the dengue transmission cycle that arise from the mosquito life cycle, the incubation period of the pathogen in the mosquito, and the incubation period of the pathogen in humans. Delay differential equations can model delayed effects because these models take into account not only the present time but also the past. For instance, in [10], the authors developed a model involving delayed (deterministic) differential equations that predicts locations of mosquito occurrence with a high accuracy, and the model realistically replicates mosquito population dynamics. The model depends on environmental drivers (temperature, precipitation, photoperiod, latitude, day of year) and human population density, and was tested with data from the *Aedes albopictus* mosquito, the most common dengue vector in Asia. By using this model, the authors analyzed the risk of dengue transmission in mainland China and concluded that temperature plays a key role in dengue outbreaks. Based on a dengue virus transmission model with maturation delay for mosquito production and seasonality, in [17] it is also found that the temperature change causes periodic oscillations of dengue fever cases.

Other usual approaches in the literature to investigate relationships between climatic factors and dengue incidence are based on regression models where overdispersion, which is often observed in dengue datasets, is taken into consideration. For instance, applying negative binomial regression models with climatic, spatial, cyclical and seasonal features as explanatory variables, ref. [18] found that precipitation, air pressure and climatic season significantly affected dengue transmission in Sri Lanka during the study period (2017–2019). In [18], all the variables were calculated with zero lags. In [19] a generalized additive model also considering a negative binomial distribution for the dengue cases (but adjusted for seasonality) was built by using climatic features with lags of 0–10 weeks and correlations were determined via Spearman's coefficient test. The model revealed that the relative humidity (with a lag of 1 week), minimum temperature (with a lag of 10 weeks) and wind (with a lag of 4 weeks) are associated with dengue cases in Asunción, Paraguay. These

authors, however, did not evaluate the fitting of the proposed model to their data. In [20], generalized linear mixed models are fitted to the number of dengue cases and allow for specific effects for different data groupings. Concretely, negative binomial regression models, with random effects related to the localization (city), the time period (year), and their interaction (city:year) are constructed to describe the associations between the dengue cases reported in 20 cities in the Brazilian state of Goiás. Spearman's correlation test is also used to identify which lags in climate factors are more correlated with cases. The authors conclude that weekly precipitation, minimum temperature, maximum temperature, and relative humidity are positively associated with dengue cases, with lags of 10,10,10, and 6 weeks, respectively.

Another way to analyse dengue data is based on stochastic models. For instance, in [21], discrete time–space stochastic SIR-SI models (susceptible-infective-recovered for human populations; susceptible-infective for vector populations) were adapted from their deterministic analogs in order to estimate the relative risk for dengue disease mapping in Malaysian states during the years 2008–2009. The authors concluded that all the states have similar patterns of expected relative risk for all epidemiological weeks.

Concerning the modelling of dengue datasets from the Dominican Republic, we highlight [22] where a generalized linear model was fitted to the cumulative reported cases for each outbreak between 2012 and 2018. In that work, the authors concluded that emerging dengue outbreaks were robust to climatological and spatiotemporal conditions, indicating that constant surveillance is necessary to prevent future outbreaks. In addition, they showed that reported dengue cases occurred mainly in the 0–15 year age group, indicating that the older age groups had higher levels of immunity. However, the effect of a time delay is not considered in this study.

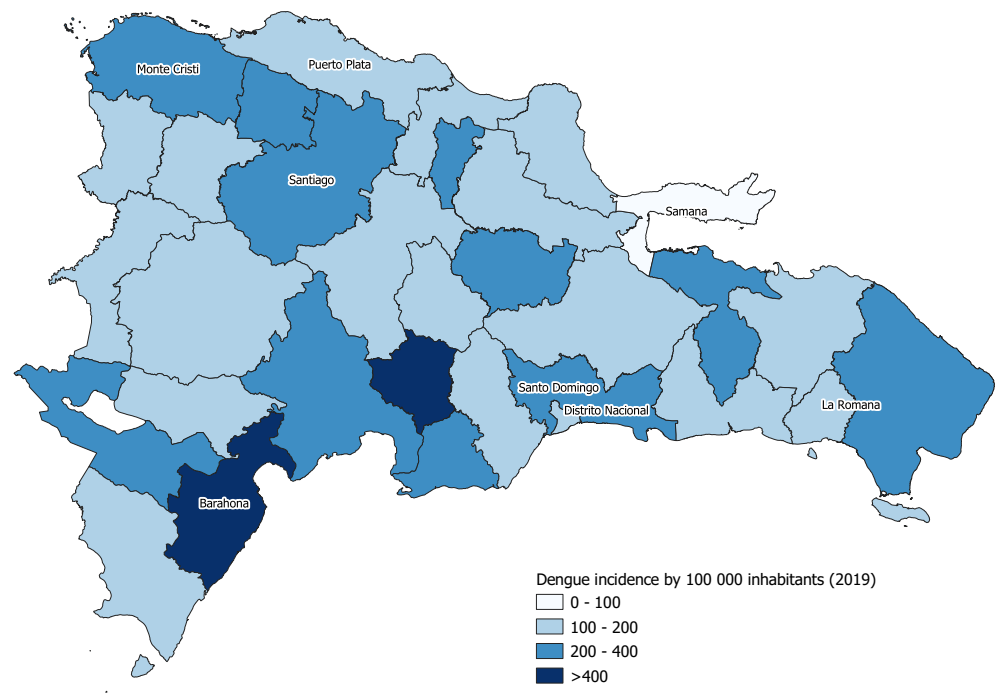
In this work, we study the dependence of the dengue incidence rate in the Dominican Republic in 2019 on delayed meteorological characteristics (temperature, rainfall, and humidity) by using gamma regression models with a normal random effects structure. The random effect is determined by geographical area (namely, the provinces) which means, in a broader sense, the modelling is conditioned to the geographical conditions of each considered area. To account for delays in transmission of dengue that arise from timing the mosquito life cycle that may be climate-dependent, we analyze relationships between dengue case data and independent meteorological variables at different times (considering lag time). For the selection of the lags, cross-correlation analysis (conditional to each province) and simple gamma regressions (one for each meteorological variable and lag) will be discussed and used to identify significant lag periods which will then be included in the final multiple regression model.

### 3. Material

#### 3.1. Geographical Area and Period of Time

This study focuses on dengue cases reported in 2019 in the Dominican Republic when a total of 20,230 dengue cases were reported corresponding to 195.3 cases per 100,000 inhabitants. The Dominican Republic is a Caribbean country on the eastern two-thirds of the Island of Hispaniola. Divided into 31 provinces plus one autonomous district (Distrito Nacional, to which we refer hereafter as one of the provinces for simplicity), the country's estimated population in 2019 was over 10.3 million people, with the metropolitan area of Santo Domingo comprising 32% of the total population [23]. The country largely experiences a tropical climate for most of the regions.

Epidemiological surveillance data for dengue and weather records were reported during 2019 for each of the 32 provinces; however, due to the lack of completeness of the data available to us, this study focuses only on eight provinces for which the total percentage of missing values was very low: Barahona, Distrito Nacional, La Romana, Monte Cristi, Puerto Plata, Samaná, Santiago, and Santo Domingo (see provinces labeled in Figure 1).



**Figure 1.** Dengue incidence per 100,000 inhabitants, in Dominican Republic in 2019. We include labels for the eight provinces included in this study.

3.2. Dependent Variable

The observed number of dengue cases officially diagnosed in hospitals of the Dominican Republic during 2019, is the dependent variable used for modelling the weekly dengue incidence rate. The data were recorded weekly and were aggregated by province. We let  $y_{ij}$  be the number of cases in province  $i$  ( $i = 1, \dots, 8$ ) reported in epidemiological week  $j$  ( $j = 1, \dots, 52$ ).

We compute five descriptive summary statistics: minimum (Min), maximum (Max), mean, standard deviation (Std.Dev), and coefficient of variation (in percentage, C.Var(%)) for the eight provinces in the study (Table 1). Throughout 2019, dengue incidence per week was, on average, highest in Santo Domingo and lowest in Samaná. Barahona had the least variability in relation to its mean comparatively among the eight provinces. For three of eight provinces, namely Barahona, Distrito Nacional, and Santo Domingo, there were dengue cases reported every week of the year.

**Table 1.** Summary statistics for dengue cases by province and by week in 2019 along with calculations of the annual dengue incidence rate (*aDIR*) and estimates of population size (Population).

	Barahona	Distrito Nacional	La Romana	Monte Cristi	Puerto Plata	Samana	Santiago	Santo Domingo
Min	4	5	0	0	0	0	0	17
Max	46	89	42	15	21	7	132	287
Mean	16.6	32.9	8.0	5.7	6.7	1.6	43.4	117.7
Std.Dev	10.9	23.7	9.9	4.9	6.4	1.7	40.3	90.2
C.Var(%)	65.9	72.0	123.5	86.4	94.6	106.7	92.89	76.6
<i>aDIR</i>	456.2	163.9	153.0	250.8	104.7	72.2	216.1	210.7
Population	189,149	1,036,494	270,166	116,605	332,386	111,217	1,038,044	2,855,892

Std.Dev, standard deviation; C.Var(%), standard deviation/mean  $\times$  100% (coefficient of variation); Population, total number of inhabitants.

Because the eight provinces vary greatly in population size, the total number of diagnosed dengue cases  $y_{ij}$  does not always give enough information about the severity of



transmission, so we calculate the total weekly incidence per 100,000 people. We calculate the incidence rate in week  $j$  in province  $i$  as

$$y_{ij}/n_i \times 100,000, \quad (1)$$

where  $n_i$  denotes the total number of inhabitants in province  $i$ . This variable will be used in our analyses in Section 4. Additionally, we analyse the annual dengue incidence rate per 100,000 inhabitants,  $aDIR_i$ , for each province  $i$ :

$$aDIR_i = \sum_j y_{ij}/n_i \times 100,000. \quad (2)$$

We include the annual dengue incidence rate per 100,000 inhabitants in Table 1 for each province studied. Additionally, Figure 1 shows the annual dengue incidence rate per 100,000 inhabitants for all 32 provinces of the Dominican Republic. Although we only have reliable meteorological data from eight provinces, it can be seen in Figure 1 that these provinces are representative of all geographic regions of the country (in the figure, provinces with name labels in the figure are those included in this study). Spatially, it is possible to observe that the disease was spread across the country. Barahona exhibited the highest value of  $aDIR$  ( $\approx 456$  cases per 100,000 inhabitants) followed by Monte Cristi ( $\approx 251$  cases per 100,000 inhabitants), which is the second-largest province by population among the eight provinces. Two provinces in the north and north east, Samaná and Puerto Plata exhibited the lowest values of  $aDIR$ , with Samaná having the lowest  $aDIR$  of fewer than 73 cases per 100,000 inhabitants.

### 3.3. Independent Variables

In this study, meteorological conditions such as temperature, rainfall, and humidity are the factors considered to influence dengue incidence. Environmental data were obtained from the Oficina Nacional de Meteorología (ONAMET), Dominican Republic, and supplemented with data from the National Aeronautics and Space Administration (NASA), of the United States of America, when some data were missing. The data include daily information on temperature (Temp; minimum, average, and maximum), precipitation (Precip; cumulative and average), relative humidity (RH; average) and daily temperature range (DTR; minimum, average, and maximum), in a total of nine variables.

#### 3.3.1. Preliminary Analysis

For each province and meteorological variable, daily data were aggregated into the 52 epidemiological weeks of 2019, giving the corresponding statistical measure per week. For instance, Temp.min is the weekly minimum of daily minimum temperatures and Temp.avg is the average temperature of the week calculated from the mean of the daily average temperatures.

For the 24 provinces not included in this study, there were missing values for many weeks (14 or more) for some independent variables. For the eight remaining provinces and the nine meteorological independent variables over the 70 weeks (=the last 18 weeks of 2018 + the 52 weeks of 2019) considered in a preliminary analysis, only about 0.7% of values were missing. Interpolations of those values were then executed by using the average of the values observed in the week before and after the occurrence of each missing value. Then, statistical measures for all the nine independent variables with respect to 2019 were calculated by province and are displayed in Table 2.

As shown in Table 2, average, maximum, and minimum temperature measurements across provinces were mostly similar. The exceptions are Monte Cristi and La Romana which both presented very low minimum values for the weekly minimum temperatures (Temp.min). For Monte Cristi, we also observed the largest coefficients of variation for both weekly minimum temperatures and maximum temperatures with C.Var(%) values of 20% and 12%, respectively. For DTR, Monte Cristi was the province whose measurements

deviated the most from the others, with low average, maximum, and minimum DTR values and higher C.Var(%) values.

**Table 2.** Summary statistics for weekly meteorological variables across the eight provinces in 2019.

Variables	Statistics	Barahona	Distrito Nacional	La Romana	Monte Cristi	Puerto Plata	Samaná	Santiago	Santo Domingo
Temp.avg	Min	24.94	26.06	22.90	20.34	24.21	24.69	23.40	26.06
	Max	29.88	30.42	28.54	26.87	29.85	29.84	29.25	30.42
	Mean	27.44	28.30	26.26	24.51	27.25	27.61	26.69	28.30
	Std.Dev	1.34	1.33	1.48	1.48	1.76	1.36	1.69	1.33
	C.Var(%)	4.87	4.70	5.65	6.05	6.47	4.93	6.33	4.70
Temp.max	Min	30.60	31.00	30.50	23.00	30.20	29.50	30.20	31.00
	Max	35.80	37.00	35.80	35.90	37.70	35.00	37.40	37.00
	Mean	32.83	33.82	32.97	27.34	33.83	32.80	33.80	33.82
	Std.Dev	1.29	1.66	1.39	3.30	2.18	1.53	1.71	1.66
	C.Var(%)	3.92	4.91	4.20	12.06	6.43	4.67	5.07	4.91
Temp.min	Min	18.00	20.10	0.00	2.20	15.60	14.00	15.00	20.10
	Max	25.00	26.00	23.20	25.00	23.60	25.00	22.00	26.00
	Mean	21.68	22.90	19.08	21.40	20.86	22.11	19.17	22.90
	Std.Dev	1.71	1.49	3.50	4.18	1.94	2.16	2.08	1.49
	C.Var(%)	7.88	6.53	18.35	19.55	9.31	9.77	10.87	6.53
DTR.avg	Min	5.99	6.94	8.99	0.00	8.54	5.26	8.76	6.94
	Max	11.30	10.50	14.10	10.20	12.57	10.39	15.47	10.50
	Mean	8.72	8.47	11.09	2.29	10.31	8.18	11.98	8.47
	Std.Dev	1.15	0.88	1.15	3.99	0.85	1.07	1.18	0.88
	C.Var(%)	13.20	10.43	10.33	174.08	8.21	13.13	9.81	10.43
DTR.max	Min	7.10	7.50	10.20	0.00	10.10	6.50	10.80	7.50
	Max	12.70	12.60	32.20	13.30	18.10	18.00	18.60	12.60
	Mean	10.44	9.96	13.18	2.97	12.25	9.90	13.91	9.96
	Std.Dev	1.29	1.04	3.00	4.98	1.40	1.84	1.58	1.04
	C.Var(%)	12.39	10.45	22.78	167.78	11.41	18.61	11.34	10.45
DTR.min	Min	3.50	3.60	4.10	0.00	5.00	-4.00	5.40	3.60
	Max	9.60	9.70	12.20	9.50	11.20	8.80	13.80	9.70
	Mean	6.82	6.97	8.99	1.55	8.50	6.67	9.77	6.97
	Std.Dev	1.41	1.03	1.53	3.04	1.20	1.85	1.82	1.03
	C.Var(%)	20.69	14.80	16.98	196.30	14.07	27.77	18.62	14.80
Precip.avg	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	14.23	11.97	21.44	15.36	21.29	21.14	20.03	11.97
	Mean	2.19	2.23	3.10	1.09	2.56	4.42	2.48	2.23
	Std.Dev	3.27	2.81	.57	2.39	3.98	4.29	3.95	2.81
	C.Var(%)	149.05	125.98	147.80	219.66	155.78	97.06	159.05	125.98
Precip.total	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	99.60	83.80	150.10	107.50	149.00	148.00	140.20	83.80
	Mean	15.00	13.73	20.98	7.63	17.63	30.69	16.94	13.73
	Std.Dev	22.65	18.23	31.32	16.86	27.76	29.84	27.48	18.23
	C.Var(%)	151.07	132.75	149.30	221.04	157.39	97.25	162.22	132.75
RH.avg	Min	64.23	72.56	76.29	60.66	75.44	77.03	69.66	72.56
	Max	83.10	84.06	90.20	77.56	90.46	89.43	88.83	84.06
	Mean	72.51	78.57	82.59	69.34	81.76	83.20	79.84	78.57
	Std.Dev	4.12	2.56	3.32	3.69	3.91	2.67	4.06	2.56
	C.Var(%)	5.69	3.26	4.02	5.33	4.78	3.21	5.08	3.26

Temp. avg., average of daily temperature observed during a week. Similar extension for the other variables: Precip., precipitation; RH, relative humidity; DTR, daily temperature range. C. Var(%), standard deviation/mean × 100% (coefficient of variation).

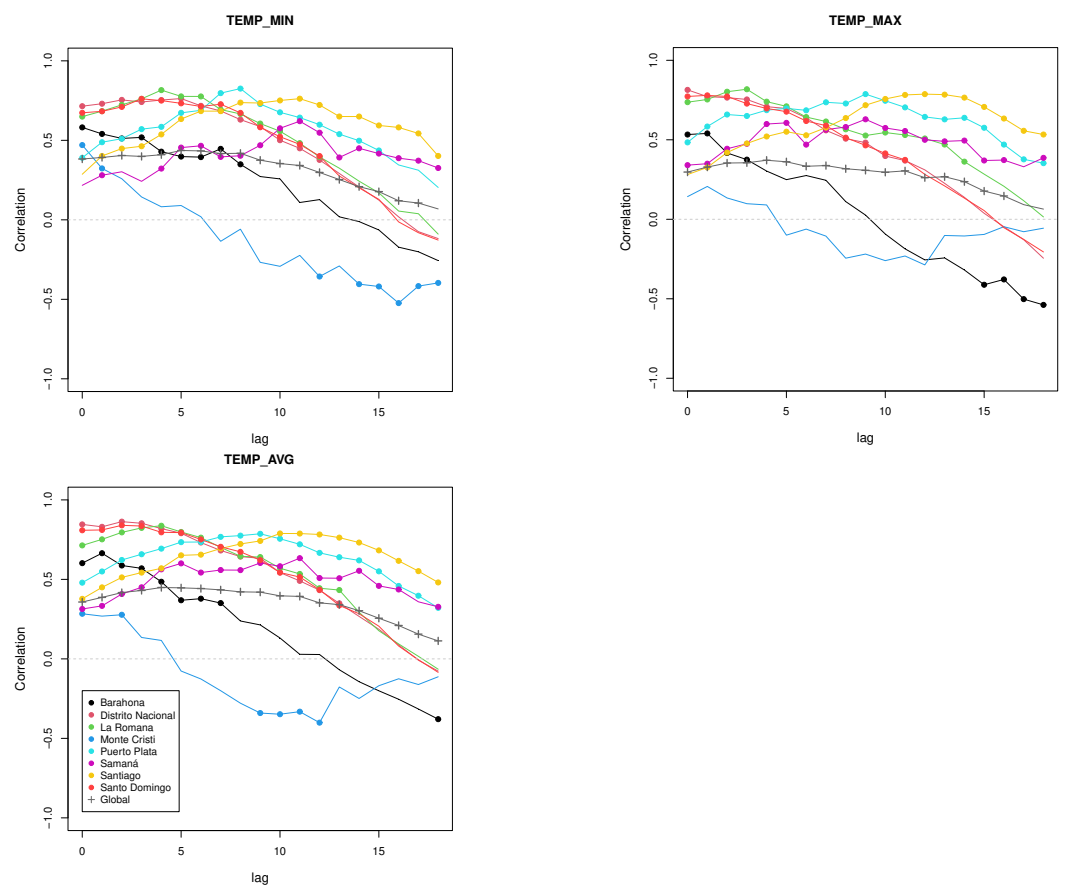
Although all eight provinces experienced weeks without rain in 2019, the maximum and average values of precipitation in 2019 tend to vary greatly among provinces. Precipitation is perhaps the factor that differs most among provinces of all the meteorological variables, with very high C.Var(%) values for both average and total weekly precipitation.

Most of the provinces present a coefficient of variation higher than 100% for both measurements with the only exception being Samaná province. Concerning the relative humidity, the weekly averages and variability therein were similar across all provinces.

### 3.3.2. Spearman’s Rank Correlation in Lags

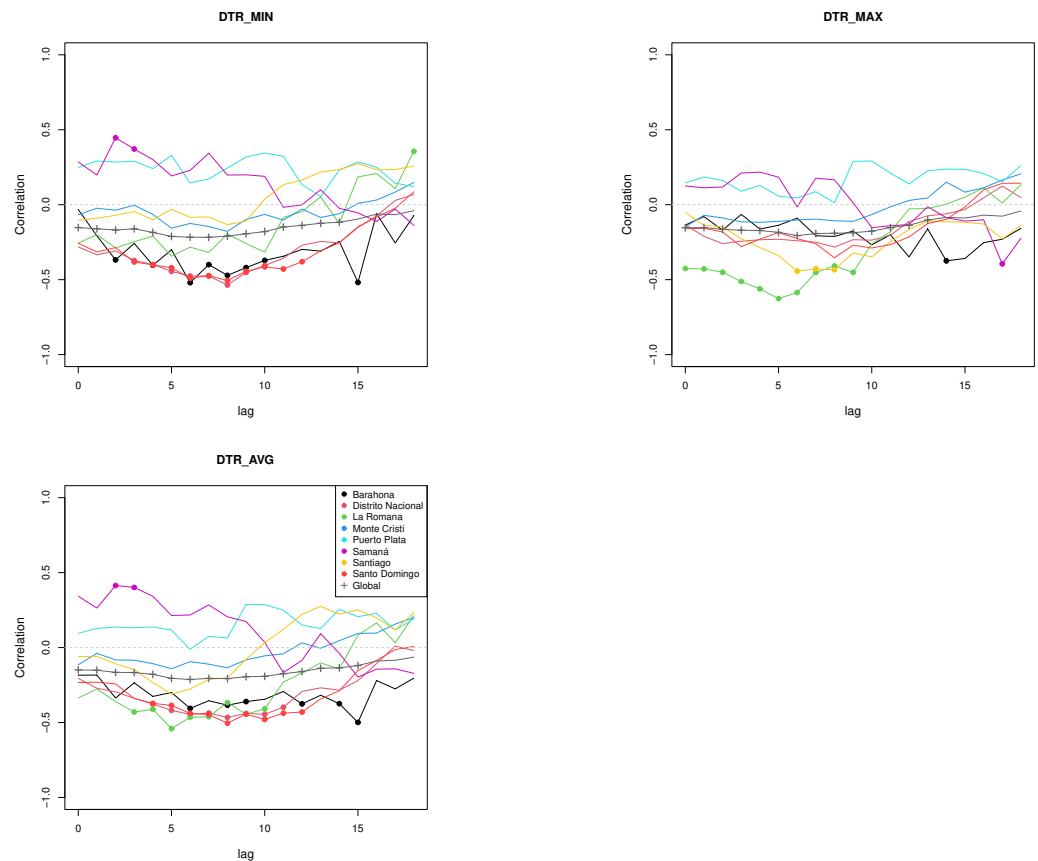
Because dengue infection relies on transmission by a mosquito vector, which in turn, along with the virus, experiences a life cycle regulated by meteorological conditions, it is important to study the time between meteorological conditions and reported cases. We consider here lags between environmental data and dengue cases. In this analysis, we include meteorological data from September 2018–December 2019 in order to consider potential impacts of weather conditions in late 2018 on transmission in early 2019.

We calculate Spearman’s rank correlation coefficients between each one of the nine meteorological variables indicated in Table 2 and the weekly dengue incidence rate with time lags of 0–18 weeks, globally (by aggregating the eight provinces:  $\sum_i y_{ij} / n_i \times 100,000$ ) and locally (by province, given by (Equation (1))). For the local analyses, because there are eight simultaneous null hypotheses for each pair lag-variable in test, the Holm procedure for multiple testing correction was applied to control the family-wise error rate at level 0.05. We show correlation coefficients for each lag for each province in Figures 2–4. In these figures, solid dots and plus points correspond to statistically significant correlations with adjusted  $p$ -values  $< 0.05$  and  $p$ -values  $< 0.05$ , respectively, with nongray colour corresponding to each of the provinces, and gray colour corresponding to the global observations.



**Figure 2.** Spearman’s rank correlation coefficient between the dengue incidence rate (1) and the minimum temperature (top left), the maximum temperature (top right) and the average temperature (bottom center) by week at lag 0–18 weeks.





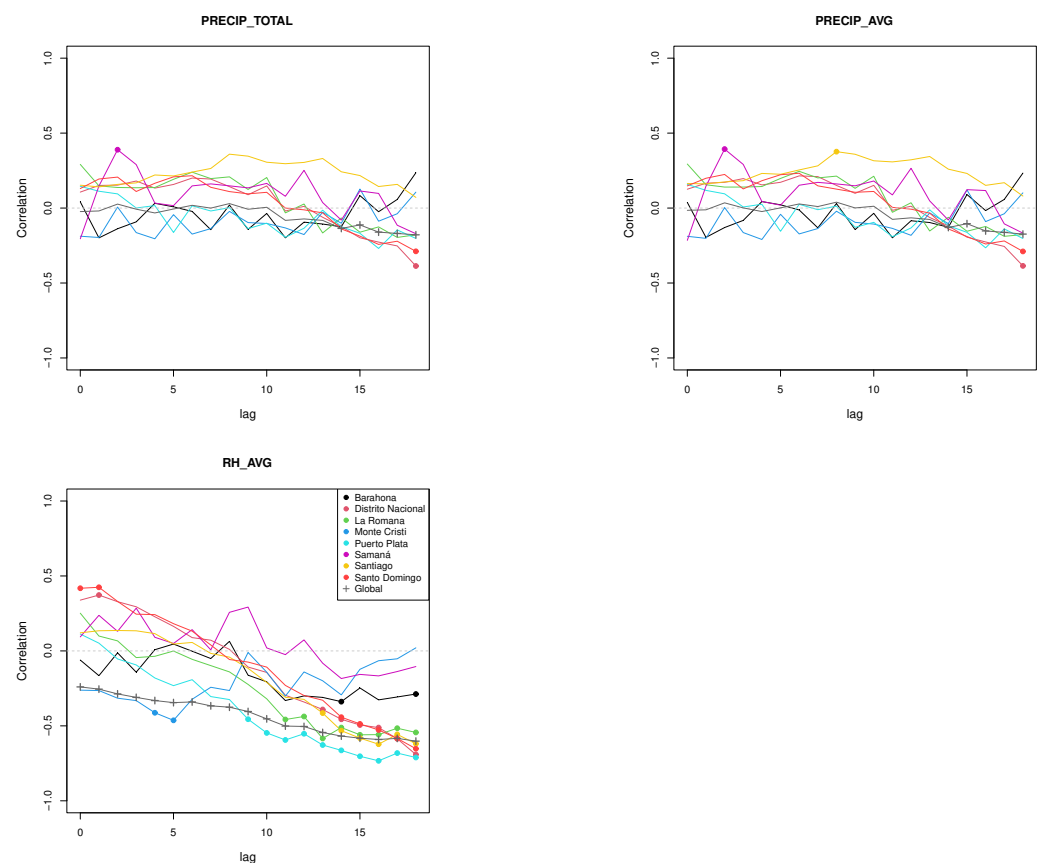
**Figure 3.** Spearman's rank correlation coefficient between the dengue incidence rate (1) and the minimum DTR (**top left**), the maximum DTR (**top right**) and the average DTR (**bottom center**) by week at lag 0–18 weeks.

These graphs show that the effect of time lag of meteorological variables on dengue incidence rate is varied among the eight provinces. For temperature variables, almost all provinces had significant positively correlated lags between dengue incidence rate and minimum, maximum, and average temperatures for many of the 19 time points we considered. Notably, Monte Cristi and Barahona were the only two provinces in which temperature variables were significantly negatively correlated with temperature. Correlation of lags with daily temperature range were much more mixed, with some provinces having negative correlations with lags in the same week that others had positive correlations in lags. However, in general, correlations in lags between dengue incidence rate and DTR were negative for most weeks studied. Regarding precipitation, very few lags were significantly correlated with dengue incidence. In fact, there are only two provinces where precipitation with lags less than 18 weeks was positively correlated with dengue cases: Samaná taking lag = 2 and Santiago taking lag = 8. Correlations in lags between dengue incidence rate and relative humidity exhibited a downward trend as the length of the lag increased: in general, correlations in smaller lags with average relative humidity were positive while correlations in larger lags were negative and often significantly so. This could be indicative of seasonal fluctuations in relative humidity.

Moreover, when we compared lags at the global level with those of provinces, we saw that correlations between climate variables, particularly temperature and DTR measurements, and dengue incidence rate in some provinces differed from the global correlations in direction and trends with increasing lag time. This potential province-specific lag effect introduces more complexity in the modelling of dengue incidence because global trends in meteorological variables may not be useful for predicting dengue at a local level, suggesting

that the inclusion of lags in variables for each province may be necessary to understand and predict dengue transmission at the province level.

To avoid increasing complexity, a singular lag per independent variable was outlined for the present work. Hence, a lag selection criterion was first established based on a positive lower lag for which there are significant associations between the independent variable, and each dependent variable for the largest number of provinces and, at the same time, preferably, for the global level. Inspecting again Figures 2–4, some lag values might visually be suggested for each climatic variable. Concretely, in Figure 2, it is observed that statistically significant associations of dengue incidence with (weekly) minimum and average temperature are only found simultaneously for all the eight provinces and globally when lag is equal to 1 and 2 weeks, respectively. However, for the remaining climatic variables, the selection of a lag value following that criterion for the eight provinces and at the global level simultaneously is more difficult. In this sense, to avoid subjectivity, further on, in Section 4, a more objective lag-selection criterion will be established.



**Figure 4.** Spearman's rank correlation coefficient between the dengue incidence rate (1) and the cumulative precipitation (**top left**), the average precipitation (**top right**) and the average relative humidity (**bottom center**) by week at lag 0–18 weeks.

#### 4. Method

For analyzing the effect of the meteorological variables on dengue incidence rate that follows a gamma distribution, gamma regressions defined by generalized linear models with fixed effects (gamma-GLM) or mixed effects (gamma-GLMM) can be adequate. Although both types of models have been applied in many studies, the latter has the advantage of being able to model clustered data structures and then incorporate within-group variability.

#### 4.1. Gamma Distribution

The gamma distribution is a probability distribution for a positive continuous random variable  $Y$  with density function given by

$$f(y; k, \theta) = \frac{1}{\Gamma(k)\theta^k} y^{k-1} \exp(-y/\theta), \quad y > 0,$$

where  $k$  and  $\theta$  are the shape parameter and the scale parameter of the distribution of  $Y$ , respectively, and  $\Gamma(k)$  is the gamma function evaluated at  $k$ . For lower values of  $k$ , the density function is right-skewed.

The expected value  $\mu$  and the variance value  $\sigma^2$  of  $Y$  are related to the shape and scale parameters in the following way:

$$\mu = k\theta \quad \text{and} \quad \sigma^2 = \mu\theta.$$

#### 4.2. Gamma Regression

Let  $\{(x_{j1}, \dots, x_{jp}, Y_j), j = 1, 2, \dots, n\}$  be a random sample defined by  $n$  independent random variables  $Y_j, j = 1, 2, \dots, n$ , such that  $Y_j | (x_{j1}, \dots, x_{jp})$  follows a gamma distribution, and consequently, its mean depends on  $p$  explanatory variables  $(x_{j1}, \dots, x_{jp})$ , i.e.,

$$\mu_j = E(Y_j | x_{j1}, \dots, x_{jp}), \quad j = 1, \dots, n.$$

In a generalized linear model with gamma responses (gamma-GLM), each mean  $\mu_j$  is described as a function of the  $p$  covariates using the link log on a linear predictor in the following form,

$$\log(\mu_j) = \beta_0 + \beta_1 x_{j1} + \dots + \beta_p x_{jp}, \quad j = 1, \dots, n, \tag{3}$$

where  $[\beta_0 \dots \beta_p]$  is a vector of  $p + 1$  unknown parameters. Because the log function is invertible, then the Gamma-GLM provides a multiplicative model to the arithmetic mean:

$$\mu_j = \exp(\beta_0 + \beta_1 x_{j1} + \dots + \beta_p x_{jp}).$$

The generalized linear mixed model with gamma responses (gamma-GLMM) is an extension of the gamma-GLM in which random effects are added to the fixed-effect parameters  $\beta_0, \beta_1, \dots, \beta_p$  in the linear predictor (3). This class of regression models is useful when there is a grouping structure of  $k$  object clusters in the set of the  $n$  data points. In such conditions, the response variable  $Y_j^{(i)}$  corresponds to the  $j$ th observation into the  $i$ th group with values of the independent variables  $x_{j1}^{(i)}, \dots, x_{jp}^{(i)}$ . Denoting  $\mathbf{b}^{(i)} = [b_1^{(i)} \dots b_q^{(i)}]$  a vector of  $q$  random effects, which are specific to the group  $i$ , the conditional response  $Y_j^{(i)} | (x_{j1}, \dots, x_{jp}, \mathbf{b}^{(i)})$  follows a gamma distribution with expected value  $\mu_{ij}$  satisfying:

$$\log(\mu_{ij}) = \beta_0^{(i)} + \beta_1^{(i)} x_{j1}^{(i)} + \dots + \beta_p^{(i)} x_{jp}^{(i)} + b_1^{(i)} z_1^{(i)} + \dots + b_q^{(i)} z_q^{(i)} \tag{4}$$

for  $j = 1, \dots, n_i, i = 1, \dots, k$ , where  $\sum_i n_i = n$ , the variables  $z_1^{(i)}, z_2^{(i)}, \dots, z_q^{(i)}$  depend on the independent variables and define the specific random effect for the  $i$ th predefined object cluster, and the vector of random effects  $\mathbf{b}^{(i)}$  is assumed to follow a  $q$ -multivariate normal distribution  $N(0, \Sigma_b)$ , where  $\Sigma_b$  is a positive definite covariance matrix independent of the cluster  $i$ .

In this work, the modelling of the dengue incidence rate in terms of the environmental variables is analyzed assuming the eight provinces as  $k = 8$  clusters in the dataset and random effect models with  $q = 1$  and  $z_1^{(i)} = 1$  in (4), and, consequently,  $\Sigma_b = \sigma^2 > 0$ .

Therefore, generalized regression models with gamma responses and a specific Y-intercept for each level of the random effects were fitted to the following models:

$$\log(\mu_{ij}) = \beta_0^{(i)} + \beta_1^{(i)} x_{j1}^{(i)} + \dots + \beta_p^{(i)} x_{jp}^{(i)} + b^{(i)}, \quad j = 1, \dots, 52, i = 1, \dots, 8. \quad (5)$$

The beta parameter  $\beta_k^{(i)}$ , for  $k = 1, \dots, p$ , can be interpreted by comparing the expected value of incidence rate  $\mu_{ij}$  at week  $j$  in the province  $i$  with that obtained when the independent  $x_k$  increases a unit and the remaining ones are not changed. In fact, from (Equation (5)) we have

$$\frac{\mu_{ij} \text{ when } x_{jk}^{(i)} + 1}{\mu_{ij} \text{ when } x_{jk}^{(i)}} = \exp(\beta_k^{(i)}), \quad j = 1, \dots, 52, i = 1, \dots, 8. \quad (6)$$

Hence, these beta parameters correspond to the logarithm of the rate ratio. Therefore, if  $\exp(\beta_k^{(i)}) > 1$ , it is expected that the dengue incidence rate, at  $k$ th epidemiological week in the province  $i$ , increases  $(\exp(\beta_k^{(i)}) - 1) \times 100\%$  for each one-unit increase in the independent variable  $x_{jk}^{(i)}$ . Otherwise, if  $\exp(\beta_k^{(i)}) < 1$ , it is expected that the dengue incidence rate, at  $k$ th epidemiological week in the province  $i$ , decreases  $(1 - \exp(\beta_k^{(i)})) \times 100\%$  for each one-unit increase in the independent variable  $x_{jk}^{(i)}$ .

#### 4.3. Gamma Fitting for the Dependent Variable

The dengue incidence rate (Equation (1)) is a continuous variable limited to the interval  $[0, +\infty)$ . It is asymmetrically distributed due to the greater presence of lower values. Given the presence of zero, the rate (Equation (1)) was (artificially) rescaled to guarantee strict positive outcomes and, consequently, a more suitable fitting of a gamma distribution. Then, an auxiliary constant equal to 0.5 was added to the numerator of (1) resulting in the adjusted rate

$$y_{ij}^* = \frac{y_{ij} + 0.5}{n_i} \times 100,000 \quad (7)$$

for the  $j$ th week and province  $i$ .

This strategy of adding an amount of 0.5 becomes relevant whenever there are no weekly records of dengue (a similar approach was also reported in other nonnormal regression models (e.g., [24])). We note that we considered alternative constants (such as 0.0001, 0.1, and 1), but numerical experiments showed that 0.5 provides good results in terms of convergence in the estimation process of the gamma-GLMM regression.

#### 4.4. Lag Selection of the Independent Variables

Because certain climatic conditions can favor the development of the mosquito and virus, and consequently, dengue transmission at a later time, effects of meteorological conditions on dengue infections were analyzed with lags. From the cross-correlation analysis of lags of 0–18 weeks performed in Section 3.3.2, it was observed that there was no clear identification of which delay week for each meteorological variable is the most important to dengue incidence rate. Thus, a more objective lag selection procedure was then established.

Concretely, the effect of each independent meteorological variable, with lags from 0 up to 5 weeks, on the adjusted incidence rate (Equation (7)) were then separately examined by using (simple) gamma-GLMM models with a single independent variable (i.e., model (5) with  $p = 1$ ). The period until 5 weeks in lags is justified, as it is a biologically plausible period of time that includes the combined time for *Ae. aegypti* egg hatching and larval development to adult mosquitoes. So, an exhaustive study with lags until 5 weeks seems to provide an adequate choice from a practical point of view.

Thus, for each meteorological variable, six simple models (simple in that there is only one independent variable and the random intercept) were considered: one for each lag-week. The lag-weeks that led to significant association ( $p$ -value  $< 0.05$ ) to the adjusted incidence rate (Equation (7)) were identified so they could be present in the final multiple model. To determine the best lag-week of each meteorological variable, four different selection criteria were then established as follows. Three criteria are based on the Akaike Information Criterion (AIC). Because all the models have the same number of parameters, comparing AIC is equivalent to comparing deviance (i.e.,  $-2 \log$  (likelihood function)). For each meteorological variable, the best lag-week was defined among the simple models with significant independent variables as

- Criterion I: the first week ( $0 \leq \text{weeks} \leq 5$ ) where a local minimum value of AIC occurred;
- Criterion II: the shortest delay ( $1 \leq \text{weeks} \leq 5$ ) where a local minimum value of AIC occurred;
- Criterion III: the delay ( $1 \leq \text{weeks} \leq 5$ ) where the global minimum value of AIC occurred;
- Criterion IV: the shortest delay ( $1 \leq \text{weeks} \leq 5$ ) where a significant association was first achieved.

In Table 3, the best lag is identified for each independent variable, in accordance with each Criterion I-IV. For RH avg., there is no significant association with the adjusted incidence rate (7), with the lowest  $p$ -value of 0.1891 calculated for a lag of 5 weeks (data not shown). For the variable DTR.min and for both rainfall variables, Precip.total and Precip.avg, the same lag was identified by all criteria: five and two delayed weeks, respectively. For the other five meteorological variables, the four criteria are not consensual about the lag set selection for the independent variables.

**Table 3.** Selection of the best lag-week based on the criteria I-IV for each independent variable.

	Criterion I		Criterion II		Criterion III		Criterion IV	
	Best Lag	AIC	Best Lag	AIC	Best Lag	AIC	Best Lag	$p$ -Value
Temp.avg	2	1767.1 **	2	1767.1 **	5	1750.2 **	0	0.0000
Temp.max	2	1847.6 **	2	1847.6 **	4	1846.7 **	0	0.0000
Temp.min	0	1876.1 **	2	1869.6 **	5	1869.3 **	0	0.0000
DTR.avg	5	1921.5 **	5	1921.5 **	5	1921.5 **	2	0.0420
DTR.max	3	1923.6 **	3	1923.6 **	5	1922.1 **	2	0.0192
DTR.min	5	1925.4 *	5	1925.4 *	5	1925.4 *	5	0.0119
Precip.avg	2	1925.1 *	2	1925.1 *	2	1925.1 *	2	0.0137
Precip.total	2	1925.6 *	2	1925.6 *	2	1925.6 *	2	0.0178
RH.avg	—	—	—	—	—	—	—	—

\*:  $0.01 \leq p$ -value  $< 0.05$ ; \*\*:  $p$ -value  $< 0.01$ .

#### 4.5. Selection of the Final Gamma-GLMMs

Given that the four criteria I-IV suggest that the lags 0, 2, 4, and 5 could be assigned to the different temperature variables and the lags 2, 3, and 5 could be assigned to the variables DTR.avg and DTR.max, multiple regression models combining these lags for the (meteorological) independent variables were constructed. One of the two precipitation variable, either Precip.average or Precip.total, was also included in the constructed models. Both precipitation variables are not simultaneously considered in the same model. For both, lag = 2 was selected in accordance with the four criteria (Table 3). The remaining variables, DTR.min and RH.avg, were also included in the models and with lag equal to 5 and 2, respectively, in accordance with the four criteria. An exhaustive comparative analysis of the constructed multiple regression models was performed to find the best multiple gamma-GLMM (Equation (5)) for estimating the adjusted incidence rate (Equation (7)). Normal and independent random intercepts  $b^{(i)}, i = 1, \dots, 8$  (defined by the eight provinces) were assumed for all the constructed models. The best-fitting final model was identified by the

lowest AIC. At the end, two multiple gamma-GLMMs, both incorporating only intercept random effects determined by provinces, were established.

#### 4.6. Validation of the Gamma-GLMMs

For checking the fitting of the gamma distribution to the observed values of the adjusted rates (7), a QQ-plot and the test based on the ratio  $V$  of two variance estimators proposed in [25] were used. For analyzing the fitting of the two final gamma GLMMs to the data, the behavior of the deviance residuals of the fitted models was examined: (i) the existence of residual patterns, globally and by province, was visually evaluated using adequate residual plots; and (ii) the normality of the deviance residuals was assessed by using QQ-plot and the Shapiro–Wilks test. To assess how well the two final gamma GLMMs performs on each province, a 95% confidence interval of the weekly average of the adjusted dengue incidence rate estimated from these two developed final models was constructed by province and verified whether the provinces contain the correspondent weekly average observed from the data.

#### 4.7. Software

All the statistical analyses were performed in the R statistical environment (R Core Team, 2020) by using the package `EnvStats` [26] for construction of gamma QQ-plot, the package `gofit` [27] for goodness-of-fit test of the gamma distribution with unknown shape and scale parameters [25], and the package `lme4` [28] for modelling of the data by gamma-GLMMs with normal random intercept. The maps of the Dominican Republic was constructed by using QGIS software.

## 5. Results

### 5.1. Selected Models

Based on the lags suggested by Criterion I-IV (Section 4.5), twelve multiple gamma-GLMMs for estimating the adjusted incidence rate (Equation (7)) in terms of meteorological regressors were constructed: models M01-M12 with lags of meteorological variables as indicated in Table 4. For the construction of these models, we considered the variable `Precip.avg` as the precipitation regressor. For six of these models, the process of estimating the parameters was not convergent (i.e., the iteration process in the optimizer via the R-function `glmer` was stopped without an optimum value for the objective function) as noted in the last column in Table 4. Among the remaining six models, the lowest AIC was achieved by the model M12, which employs as explanatory variables the weekly average and minimum temperatures and average DTR with a lag of 5 weeks, the weekly maximum temperature with a lag of 4 weeks, and the average weekly precipitation with a lag of 2 weeks. M12 also included the maximum and minimum DTR with a lag of 5 weeks, but these variables were not significant to predictions.

In fact, the variable `DTR.max` was not a significant predictor in any of the previous models, so it was removed, and we replicated the procedure as before. With `DTR.max` removed, a total of eight models were developed, M13-M20 as indicated in Table 4. Among the convergent models, the lowest AIC was achieved to the model M20, which included all of the same predictors as M12 excepting `DTR.max`.

For both models, M12 and M20, the same set of independent variables was obtained as statistically significant: `Temp.avg`, `Temp.max`, `Temp.min`, `DTR.avg`, `DTR.min` and `Precip.avg` with lag equals to 5, 4, 5, 5, 5, and 2 weeks, respectively. Finally, because the variable `DTR.min` (with lag = 5 weeks) is not statistically significantly in the model, it was removed at this stage to obtain the final model (M21 in Table 4) which reached a slightly lower AIC value than the previous ones.

In this final model, temperature and DTR variables have larger lags. Therefore, this model could represent a long-term alarm based on both temperature and DTR conditions. Therefore, it will be called a longer-term model. Given this result, we aimed to next answer the question: could these same variables predict adjusted dengue incidence rate in a shorter



time? Consequently, an extra model (M22 in Table 4), called a shorter-term model, with all these variables with a delay of 2 weeks was analyzed. All variables with a delay of 2 weeks were significant predictors; however, the AIC for M22 was not lower than that of M21, indicating that it is not a better model overall.

**Table 4.** Lag for the meteorological regressors in 22 multiple gamma-GLMMs constructed for fitting the dengue incidence rate. Statistical significance of the regressors are identified.

Model	Temperature			DTR			Precipitation	AIC
	avg.	max.	min.	avg.	max.	min.	avg.	
M01	0	0	0	2	2	5	2	nc
M02	0 (0.0003)	0	0	5 (0.0659)	3	5	2	1806.5
M03	0	0	0	5	5	5	2	nc
M04	2 (<0.0001)	2 (0.0905)	0	2 (0.0174)	2	5 (0.0744)	2 (<0.0001)	1738.9
M05	2 (<0.0001)	2 (0.0309)	0	5 (0.0021)	3	5	2 (<0.0001)	1729.5
M06	2	2	0	5	5	5	2	nc
M07	2	2	2	2	2	5	2	nc
M08	2	2	2	5 (0.0026)	3	5	2 (<0.0001)	1728.1
M09	2 (<0.0001)	2	2	5 (0.0038)	5	5 (0.0950)	2 (<0.0001)	1728.6
M10	5	4	5	2	2	5	2	nc
M11	5	4	5	5	3	5	2	nc
M12	5 (<0.0001)	4 (0.0006)	5 (0.0153)	5 (0.0076)	5	5	2 (0.0296)	1699.3
M13	0 (0.0002)	0	0	2 (0.0086)	–	5	2 (0.0001)	1804.8
M14	0 (0.0002)	0	0	5 (0.0086)	–	5	2 (0.0001)	1804.8
M15	2	2	0	2	–	5	2	nc
M16	2 (<0.0001)	2 (0.0523)	0	5 (<0.0001)	–	5	2 (<0.0001)	1728.6
M17	2	2	2	2	–	5	2	nc
M18	2 (<0.0001)	2	2	5 (<0.0001)	–	5	2 (<0.0001)	1727.2
M19	5	4	5	2	–	5	2	nc
M20	5 (<0.0001)	4 (0.0006)	5 (0.0120)	5 (<0.0001)	–	5	2 (0.0250)	1697.3
M21	5 (<0.0001)	4 (0.0004)	5 (0.0120)	5 (<0.0001)	–	–	2 (0.0258)	1695.5
M22	2 (<0.0001)	2 (0.1260)	2 (<0.0001)	2 (<0.00005)	–	–	2 (0.0258)	1736.8

The *p*-value is indicated in parentheses; nc means the optimization routine was non-convergent.

We note that the procedure described above was replicated substituting the variable Precip.avg by Precip.total as a precipitation regressor. Under this condition, more nonconvergent models emerged, and slightly higher values of AIC were in general produced for the convergent homologous models (data not shown). We also note that when the random intercepts  $b^{(i)}$  are not included in the multiple model (5) (i.e., where the approach GLM is used for modelling of the adjusted dengue incidence rate (Equation (7))), substantially higher values of AIC would be obtained for the correspondent GLMs, namely AIC = 1923.7 and AIC = 1936.8 for the fitted longer-term model and shorter-term model, respectively, justifying the modelling of the dengue data set by GLMM than by GLM.

As indicated in Table 4, in the longer-term model (M21), the weekly dengue incidence rate is described in of terms the average and the minimum values of the temperature both

delayed by 5 weeks (Temp.avg5 and Temp.min5), the maximum value of the temperature delayed by 4 weeks (Temp.max4), and the average value of precipitation delayed by 2 weeks (Precip.avg2). In the *shorter-term* model (M22), the weekly dengue incidence rate is described by effects of those five meteorological features all with a delay of 2 weeks (Temp.avg2, Temp.max2, Temp.min2, DTR.avg2, and Precip.avg2).

Formally, these two models are defined as follows,

$$\log(\mu_{ij}) = \beta_0 + \beta_1 \text{Temp.avg5}_j + \beta_2 \text{Temp.max4}_j + \beta_3 \text{Temp.min5}_j + \beta_4 \text{DTR.avg5}_j + \beta_5 \text{Precip.avg2}_j + \text{Province}^{(i)}$$

and

$$\log(\mu_{ij}) = \beta_0 + \beta_1 \text{Temp.avg2}_j + \beta_2 \text{Temp.max2}_j + \beta_3 \text{Temp.min2}_j + \beta_4 \text{DTR.avg2}_j + \beta_5 \text{Precip.avg2}_j + \text{Province}^{(i)}$$

respectively, for each province  $i = 1, \dots, 8$ , and meteorological conditions summarized in epidemiological week  $j = 1, \dots, 52$  of 2019. The variable  $\text{Province}^{(i)}$  is assumed to be normally distributed with zero mean and constant variance  $\sigma_i^2$  in the estimation process of the regressor coefficients in each model and represents the random effect specific (Y-intercept) to the  $i$ th province.

### 5.2. Validation

The histogram and the QQ-plot presented in Figure 5 suggest that the empirical right-skewed distribution of the adjusted rates (Equation (7)) for the set of the eight provinces is close to a gamma distribution (with  $\mu \approx 4.29$  and  $\sigma \approx 4.09$ ). There was no significant evidence that a gamma distribution did not provide an adequate fit ( $V = 0.51294$ ,  $p$ -value = 0.7168).

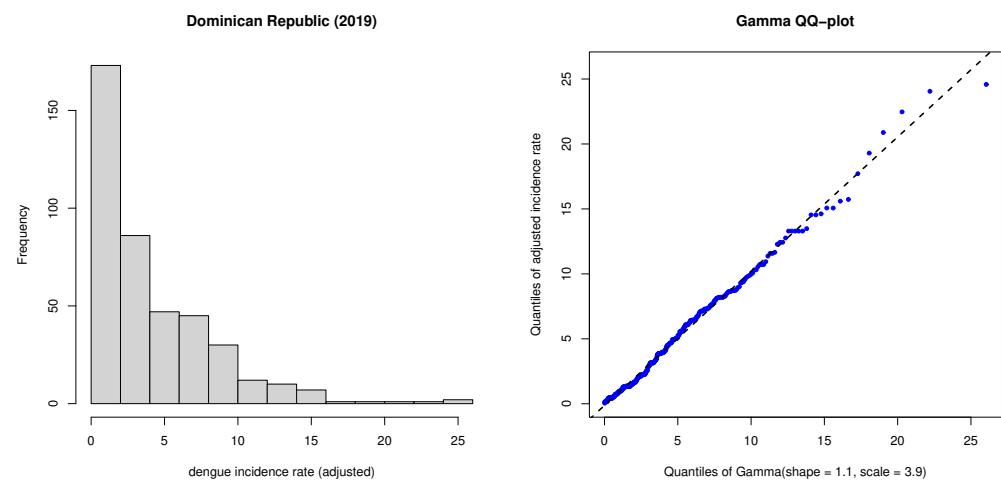
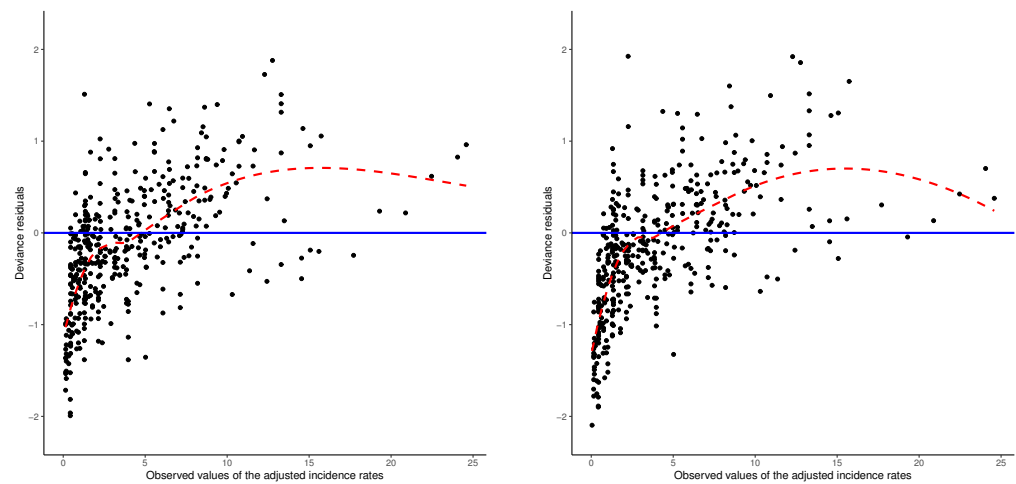


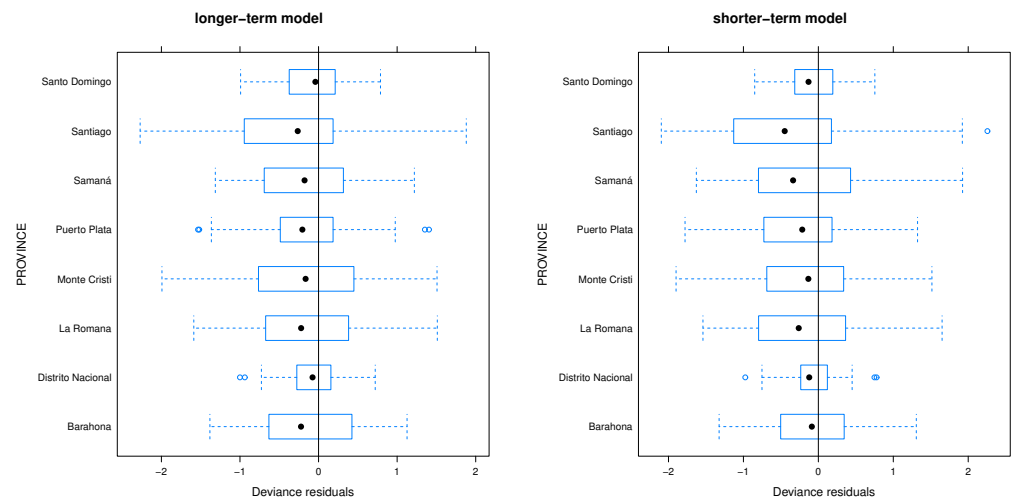
Figure 5. Histogram and gamma QQ-plot for the adjusted dengue incidence rate.

In Figure 6, for both models, it is observed that almost all of the deviance residuals vary between  $-2$  and  $2$  and there is a higher spread of points for higher observed values of the incidence rate (7). When the observed incidence rate is close to zero, there are many negative residuals suggesting that both models tend to predict higher incidence rates than the observed rates. For higher observed values (e.g., for weekly incidence rate between approximately 6 and 15 per 100,000 inhabitants), both the smoothed average curves (red lines in the graphs) tend to increase, indicating that the values predicted by both models will be lower than the observed. Nevertheless, for situations with the highest incidence rates, the fitted curve is closer to zero in the shorter-term model. This suggests that meteorological conditions of temperature, DTR and precipitation of 2 weeks earlier tend to provide better predictions for dengue incidence when outbreaks are larger than than those predictions using longer delays.



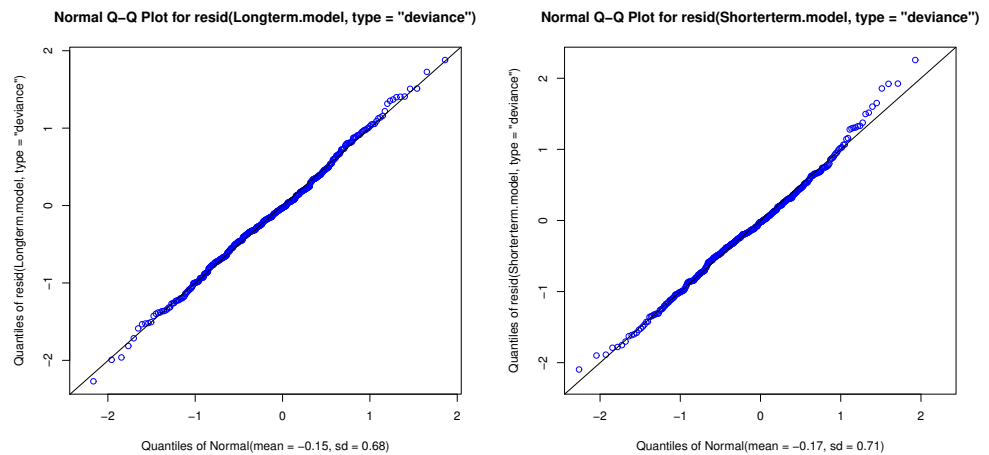
**Figure 6.** Deviance residuals versus observed values of the longer-term model (on the left) and the shorter-term model (on the right) with smooth loess curve (in red).

In Figure 7, the comparative boxplots of the deviance residuals for the eight provinces show that (i) the deviance residuals only exceeds the interval  $[-2, 2]$  in Santiago, and only slightly; (ii) there are outliers, suggesting that there are a few weeks when the model estimates of the incidence rates could be atypical (in Distrito Nacional, Puerto Plata and Santiago); and (iii) the variability in two provinces, Santo Domingo and Distrito Nacional, seems to be lower than in other provinces. Although these results indicate the models do not always predict the true incidence rates in some weeks and some provinces, both the models fit relatively well to the data.



**Figure 7.** Comparative boxplots for deviance residuals across the eight provinces.

In Figure 8, the good alignment of the points with the diagonal line in both the QQ-plots suggests a normal distribution to the deviance residuals for both models. From the Shapiro–Wilks test, there was no significant evidence that the distributions of deviance residuals of both models were nonnormal ( $p$ -value = 0.818 for longer-term model and  $p$ -value = 0.136 for the shorter-term model).



**Figure 8.** QQ-plots for the deviance residual related to the longer-term model (on the left) and the shorter-term model (on the right).

5.3. Interpretation

The estimates of the fixed effects and the effect variance of the two fitted models are displayed in Table 5. Only a single covariate coefficient (Temp.max2 for the shorter-term model) was not statistically significantly different from zero at a 5% significance level. Therefore, associations between each significant meteorological variable and the dengue incidence rate for the eight provinces of Dominican Republic can be then described assuming that the remaining variables are fixed. Variations in the daily average temperature (Temp.avg) have the greatest effect on the dependent variable (dengue incidence rate) with an increase of 1 °C leading to an increase in the dengue incidence rate by 52.4% ( $\exp(0.4212) = 1.5238$ ) 2 weeks later and 44.4% ( $\exp(0.3674) = 1.4440$ ) 5 weeks later. Although the same increase of the maximum temperature (Temp.max) increases the dengue incidence rate by 3.5% ( $\exp(0.0341) = 1.0347$ ) 2 weeks later and 13% ( $\exp(0.1218) = 1.1295$ ) 4 weeks later, the minimum temperature (Temp.min) reduces the rate of reported cases by 5.0% ( $\exp(-0.0510) = 0.9503$ ) 2 weeks later and 6.0% ( $\exp(-0.0618) = 0.9401$ ) 5 weeks later. If the average daily temperature range (DTR.avg) is 1 °C higher, then a decrease of 18.3% ( $\exp(-0.2025) = 0.8167$ ) and 11.5% ( $\exp(-0.1224) = 0.8848$ ) in dengue incidence rate is observed 5 and 2 weeks later, respectively. A 1-mm increase in the weekly average precipitation (Precip.avg) triggers an increase in the dengue incidence rate of 2.0% ( $\exp(0.0210) = 1.0212$ ) and 4.5% ( $\exp(0.0436) = 1.0446$ ) 2 weeks later for the longer-term and shorter-term models, respectively.

**Table 5.** Parameter estimates of two fitted gamma-GLMMs for modelling the adjusted incidence dengue rate.

Parameter	Lag	Longer-Term Model			Shorter-Term Model			
		Beta	Std. Error	p-Value	Lag	Beta	Std. Error	p-Value
(Intercept)		−9.7579	0.7087	0.0000		−9.3166	0.7250	0.0000
Temp.avg	5	0.3674	0.0526	0.0000	2	0.4212	0.0676	0.0000
Temp.max	4	0.1218	0.0343	0.0004	2	0.0341	0.0476	0.4743
Temp.min	5	−0.0618	0.0247	0.0122	2	−0.0510	0.0247	0.0388
DTR.avg	5	−0.2025	0.0249	0.0000	2	−0.1224	0.0296	0.0000
Precip.avg	2	0.0210	0.0094	0.0258	2	0.0436	0.0101	0.0000

The variances of the random effects,  $\sigma^2_{province}$ , were estimated to be equal to 0.2363 and 0.2764 for the longer-term and shorter-term models, respectively. This result indicates slightly lower variability among the eight provinces for the Y-intercept of the fitted model when the meteorological variables are considered with more delays.

In Table 6, we show the observed adjusted dengue incidence rate along with 95% confidence intervals in the weekly average adjusted dengue incidence rate estimated by using both longer-term and shorter-term models across the eight provinces under study. The observed values fall within the estimated 95% confidence interval in all cases except for in Barahona for the longer-term model. Consequently, the longer-term model overestimates the dengue incidence rate in Barahona.

**Table 6.** Weekly average of the observed value and 95% confidence intervals (CI) for the weekly average of the estimated value from the two fitted gamma-GLMMs across the 52 weeks of 2019 for the adjusted dengue incidence rate for each province.

Provinces	Longer-Term Model			Shorter-Term Model		
	Lower CI	Upper CI	Observed	Lower CI	Upper CI	Observed
Barahona	9.377	13.160	9.038	8.549	11.732	9.038
Distrito Nacional	2.767	3.826	3.221	2.685	3.597	3.221
La Romana	2.257	3.193	3.153	2.289	3.177	3.153
Monte Cristi	4.788	6.356	5.278	4.473	5.869	5.278
Puerto Plata	1.726	2.472	2.117	1.788	2.518	2.170
Samaná	1.709	2.333	1.850	1.837	2.459	1.850
Santiago	3.160	4.504	4.233	3.764	5.406	4.233
Santo Domingo	3.357	4.642	4.140	3.274	4.386	4.140

In Table 7, estimates of the random effects are presented. Comparing longer-term and shorter-term models, we observe very similar negative estimates of random intercepts between models among four provinces: Santo Domingos, Puerto Plata, Distrito Nacional and Samaná. This indicates that, based on meteorological variables with shorter and longer delays, with lags as described in Table 5 for both models, lower expected values for the adjusted dengue incidence rate are predicted for these four provinces, with Samaná presenting the lowest one. Among the other four provinces, which have positive estimates, similar values between the two models are only observed for Santiago’s province. The highest estimate of the random intercept of the longer-term model occurs for the province of Barahona and for the shorter-term model the highest estimate occurs for Monte Cristi. Therefore, although the adjusted dengue incidence rate based on 2 week-lag meteorological variables is expected to be higher in Monte Cristi, meteorological variables with a longer delay, with lags as indicated in the longer-term model will lead to a higher estimated rate in Barahona.

**Table 7.** Estimated intercept random effects for each of the eight provinces for the two fitted gamma-GLMMs.

Provinces	Random Effects	
	Longer-Term Model	Shorter-Term Model
Barahona	0.977	0.873
Distrito Nacional	−0.648	−0.646
La Romana	0.240	0.136
Monte Cristi	0.747	0.908
Puerto Plata	−0.550	−0.535
Samaná	−0.841	−0.851
Santiago	0.536	0.567
Santo Domingo	−0.454	−0.448

### 6. Conclusions

Dengue outbreaks are a consequence of complex interactions among multiple factors. In particular, dengue disease depends on the development of mosquitoes through a four-stage life cycle that is heavily influenced by environmental conditions [29]. This

implies that the current number of cases can be influenced by past conditions that impact the mosquito life cycle. By using consistent methods for fitting gamma-GLMM models, we have analysed the effect of meteorological conditions with lags on the incidence of dengue conditioned to the human population density of eight provinces of the Dominican Republic. We defined two relationships in terms of province-specific effects and different statistics for meteorological variables related to temperature (average, maximum, minimum), daily temperature range (average), and precipitation (average) to explain the dengue incidence rate. Although one model provides estimates of dengue incidence by using meteorological variables in the short term (2 weeks), the other describes dengue incidence in terms of meteorological conditions reported after passage of more time. Our results showed a significant effect from temperatures with delay of 2, 4, and 5 weeks, from daily temperature range with delay of 2 and 5 weeks and from precipitation with delay of 2 weeks. Additionally, variations in average temperature (Temp.avg) have the greatest effect on dengue cases. These results are in agreement with similar studies that found significant risk of dengue when considering lags in climate variables of 2–5 weeks [3,4]. Our findings provide a better understanding of the relationships between meteorological conditions and weekly trends in dengue cases during the outbreak that occurred in the Dominican Republic in 2019.

Different geographical and spatial locations may have local effects that lead to different dengue models. We note that the focus of this work is on the effects of meteorological variables and the modelling of the dengue cases is conditioned to the population size of each province. But other influencers of dengue transmission are likely. We included province-specific random effects to account for some of the variability that could be produced by these influencers. In particular, geographical and sociopolitical features of each province could play a role in dengue transmission; however, these features are not considered as independent variables in this study. Analysing the variances of random effects, it is possible to conclude that among the eight provinces studied there is lower variability. But the distribution of values for random effects (Table 7) suggest that it is important to question whether other factors could be considered to improve model predictions. Our results highlight that dengue prediction models developed at local scales are important to understanding the risk of dengue because conditions at a higher level (such as at the national level), may not be useful for predicting dengue cases that have high heterogeneity driven by different geographic, climatic, sociodemographic, or other factors.

The analysis of the deviance residuals shows that, overall, both the models fit relatively well to the data. A normal distribution is fitted to the deviance residuals; however, there is a variability of the incidence rate within the provinces which suggest that both models might be improved by the addition of more random effects. The diagnostic model showed that there is variability in the incidence rate between provinces. Using the same lag for all provinces could not be the best choice; instead, it is important to investigate whether the selected lags in the meteorological variables are province-specific. For this reason, further studies are needed to develop a better understanding of variability in dengue incidence rates.

The lack of reliable data for a long period of time was a limitation of this work. There is a lack of long-term spatiotemporal and climate data for dengue incidence in the Dominican Republic. In spite of having data for 32 provinces, we only had reliable data for eight of them. Even considering eight provinces, it was necessary to supplement climate data collected in the Dominican Republic with data collected from other sources (NASA) to reduce missing values in the database. For future works, complete climate data from other provinces in the Dominican Republic is necessary. Having data from more provinces would allow us to assess how well the two final models perform in other provinces. Moreover, having data from more provinces and years could allow us to establish a better final model by using another strategies like cross-validation by splitting the data into two sets: a training data set for model-selection and a testing data set for inference. In order to develop better models aimed at understanding the intensification of dengue transmission in the Dominican Republic and to develop reliable warning systems for predicting future dengue



incidence, it is important to gather more long-term data and to build robust systems for continuous collection of this data.

This research contributes to developing a better understanding of the dynamics of dengue and their relationship with climatological variables in the Dominican Republic, a tropical country where, despite minor differences in climate across the country, dengue incidence can vary greatly. This paper has practical implications for preparing vector control and public health departments by providing potential warning indicators for dengue outbreaks, which will in turn contribute to the development of comprehensive dengue management programs.

**Author Contributions:** Conceptualization, A.F., H.S.R. and N.M.; methodology, A.F., H.S.R. and N.M.; software, A.F., H.S.R. and A.I.; formal analysis, A.F., H.S.R. and N.M. and M.A.R.; validation, M.A.R., D.H. and M.C.-H.; data curation, A.I., M.A.R., D.H. and M.C.-H.; writing—original draft preparation, A.F., H.S.R., N.M. and M.A.R.; writing—review and editing, A.F., H.S.R., N.M., A.I., M.A.R., D.H. and M.C.-H.; supervision, A.F. and H.S.R.; project administration, A.F. and H.S.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** Fund for Innovation and Scientific and Technological Development—Ministry of Higher Education, Science and Technology of the Dominican Republic. Work supported by Portuguese funds through the CIDMA—Center for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology (FCT-Fundação para a Ciência e a Tecnologia), within project UIDB/04106/2020.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Dengue and Severe Dengue, World Health Organization. 2021. Available online: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue> (accessed on 22 March 2022).
2. Ten Threats to Global Health in 2019, World Health Organization. Available online: <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019> (accessed on 9 April 2022)
3. Iutis, A.; Rodrigues, H.S.; Freitas, A.; Martins, N. A preliminary analysis of weekly dengue incidence rate in the Dominican Republic during 2019. *J. Stat. Health Decis.* **2021**, *3*, 6–9.
4. Robert, M.A.; Rodrigues, H.; Sofia, H.D.; Donado Campos, J.M.; Morilla, F.; Aguila Mejia, J.; Guardado, M.E.; Skews, R.; Colome-Hidalgo, M. Spatiotemporal and Meteorological Trends in Dengue Transmission in the Dominican Republic, 2015–2019. Available online: <https://medrxiv.org/cgi/content/short/2023.01.05.23284205v1> (accessed on 10 January 2023).
5. Brady, O.J.; Johansson, M.; A Guerra, C.; Bhatt, S.; Golding, N.; Pigott, D.M.; Delatte, H.; Grech, M.G.; Leishnam, P.T.; Maciel-De-Freitas, R.; et al. Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasites Vectors* **2013**, *6*, 351. [CrossRef] [PubMed]
6. Reinhold, J.M.; Lazzari, C.R.; Lahondère, C. Effects of the Environmental Temperature on *Aedes aegypti* and *Aedes albopictus* Mosquitoes: A Review. *Insects* **2018**, *9*, 158. [CrossRef] [PubMed]
7. Halstead, S.B. Dengue. *Lancet* **2007**, *70*, 1644–1652. [CrossRef]
8. Srisuphanunt, M.; Puttaruk, P.; Kooltheat, N.; Katzenmeier, G.; Wilairatana, P. Prognostic Indicators for the Early Prediction of Severe Dengue Infection: A Retrospective Study in a University Hospital in Thailand. *Trop. Med. Infect. Dis.* **2022**, *31*, 162. [CrossRef]
9. Patanarapelert, K.; Tang, I.M. *Effect of Time Delay on the Transmission of Dengue Fever*; World Academy of Science, Engineering and Technology: Chicago, IL, USA, 2007.
10. Metelmann, S.; Liu, X.; Lu, L.; Caminade, C.; Liu, K.; Cao, L.; Medlock, J.M.; Baylis, M.; Morse, A.P.; Liu, Q. Assessing the suitability for *Aedes albopictus* and dengue transmission risk in China with a delay differential equation model. *PLoS Neglected Trop. Dis.* **2021**. [CrossRef]
11. Rodrigues, H.S.; Monteiro, M.T.T.; Torres, D.F.M. Vaccination Models and Optimal Control Strategies to Dengue. *Math. Biosci.* **2014**, *247*, 1–12. [CrossRef]
12. Gan, S.J.; Leong, Y.Q.; Bin Barhanuddin, M.F.H.; Wong, S.T.; Wong, S.F.; Mak, J.W.; Ahmad, R.B. Dengue fever and insecticide resistance in *Aedes* mosquitoes in Southeast Asia: A review. *Parasites Vectors* **2021**, *14*, 315. [CrossRef]
13. Jayawickreme, K.P.; Jayaweera, D.K.; Weerasinghe, S.; Warapitiya, D.; Subasinghe, S. A study on knowledge, attitudes and practices regarding dengue fever, its prevention and management among dengue patients presenting to a tertiary care hospital in Sri Lanka. *BMC Infect. Dis.* **2021**, *21*, 981. [CrossRef]

14. Brito da Cruz, A.M.C.; Rodrigues, H.S. Economic Burden of Personal Protective Strategies for Dengue Disease: An Optimal Control Approach. In *Optimization, Learning Algorithms and Applications; OL2A 2021; Communications in Computer and Information Science; Springer: Cham, Switzerland, 2021; Volume 1488.*
15. Brito da Cruz, A.M.C.; Rodrigues, H.S. Personal protective strategies for dengue disease: Simulations in two coexisting virus serotypes scenarios. *Math. Comput. Simul.* **2021**, *188*, 254–267. [[CrossRef](#)]
16. Sebastião, C.S.; Neto, Z.; Jandondo, D.; Mirandela, M.; Morais, J.; Brito, M. Dengue virus among HIV-infected pregnant women attending antenatal care in Luanda, Angola: An emerging public health concern. *Sci. Afr.* **2022**, *17*, e01356. [[CrossRef](#)]
17. Song, H.; Tian, D.; Shan, C. Modeling the effect of temperature on dengue virus transmission with periodic delay differential equations. *Math. Biosci. Eng.* **2020**, *17*, 4147–4164. [[CrossRef](#)] [[PubMed](#)]
18. Faruk, M.O.; Jannat, S.N.; Rahman, M.S. Impact of environmental factors on the spread of dengue fever in Sri Lanka. *Int. J. Environ. Sci. Technol.* **2022**, *19*, 10637–10648. [[CrossRef](#)]
19. Gómez-Gómez, R.E.; Kim, J.; Hong, K.; Jang, J.Y.L.; Kisiju, T.L.; Kim, S.; Chun, B.C. Association between Climate Factor and Dengue Fever in Asuncion, Paraguay: A Generalized Additive Model. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12192. [[CrossRef](#)] [[PubMed](#)]
20. Oliveira, A.N.; Menezes, R.; Faria, S.; Afonso, P. Mixed-effects modelling for crossed and nested data: An analysis of dengue fever in the state of Goiás, Brazil. *J. Appl. Stat.* **2020**, *47*, 2912–2926. [[CrossRef](#)]
21. Samat, N.A.; Percy, D.F. Vector-borne infectious disease mapping with stochastic difference equations: An analysis of dengue disease in Malaysia. *J. Appl. Stat.* **2012**, *39*, 2029–2046. [[CrossRef](#)]
22. Petrone, M.E.; Earnest, R.; Lourenço, J.; Kraemer, M.U.; Paulino-Ramirez, R.; Grubaugh, N.D.; Tapia, L. Asynchronicity of endemic and emerging mosquito-borne disease outbreaks in the Dominican Republic. *Nat. Commun.* **2021**, *12*, 151. [[CrossRef](#)]
23. Oficinal Nacional de Estadística. Available online: <https://www.one.gob.do/datos-y-estadisticas/temas/estadisticas-demograficas/estimaciones-y-proyecciones-demograficas/> (accessed on 15 February 2022).
24. Sciandra, M.; Spera, I. A model-based approach to Spotify data analysis: A Beta GLMM. *J. Appl. Stat.* **2022**, *49*, 214–229. [[CrossRef](#)]
25. Villasenor, J.A.; Gonzalez-Estrada, E. A variance ratio test of fit for Gamma distributions. *Stat. Probab. Lett.* **2015**, *96*, 281–286. [[CrossRef](#)]
26. Millard, S.P. *EnvStats: An R Package for Environmental Statistics*; Springer: New York, NY, USA, 2013.
27. Gonzalez-Estrada, E.; Villasenor-Alva, J.A. *Goft: Tests of Fit for Some Probability Distributions*. R Package Version 1.3.6. 2020. Available online: <https://CRAN.R-project.org/package=goft> (accessed on 5 December 2022)
28. Bates, D.; Maechler, M.; Bolker, B.; Walker, S. Linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [[CrossRef](#)]
29. Navarro Valencia, V.; Díaz, Y.; Pascale, J.M.; Boni, M.F.; Sanchez-Galan, J.E. Assessing the Effect of Climate Variables on the Incidence of Dengue Cases in the Metropolitan Region of Panama City. *Int. J. Environ. Res. Public Health* **2021**, *18*, 12108. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.