Research article

Fitting the return period of dengue fever epidemic in Guangdong province of China

Siqing Zeng [a],[*], Jianpeng Xiao [a], Fen Yang [b], Jiya Dai [b], Meng Zhang [b], Haojie Zhong [b]

[a] Guangdong Provincial Institute of Public Health, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, 511430, Guangdong, China
[b] Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, 511430, Guangdong, China

A B S T R A C T

Objectives: The prevention and control of dengue fever (DF) has been a major public health issue in Guangdong (GD) province, China. This study aims to analyze the return period (RP) and the return level (RL) of DF epidemic in GD, to help the formulation of prevention and control plan.
Methods: Three models, namely Lognormal distribution (Lognor D.), normal distribution (Norm D.), and generalized logistic distribution (GLD) were selected to fit the annual number of indigenous DF cases in GD from 1978 to 2021. The coefficient of determination (R^2), the root mean squared error (RMSE), and the Akaike information criterion (AIC) were used to evaluate the goodness of fit. We predicted the RP of 45130 historical maximum cases that occurred in 2014 and the RP of 4884 peak cases that occurred in 2019 over the 5 years up to 2021.
Results: Fitting through the three models, the R^2 was 0.98, 0.98, and 0.96, respectively. The predicted RLs of the annual DF case number were between 297 and 43234, 297 and 43233, 362 and 41868 for the RPs of 2–45 years. The predicted RPs of DF outbreaks exceeding the historical maximum were 43, 43, and 44 years, and the RPs of DF epidemic exceeding the peak in 2019 were 7, 7, and 8 years, respectively. Therefore, we predicted that GD would experience a DF outbreak beyond the historical maximum in the next 35 or 36 years from 2022. And in the next 4 or 5 years from 2022, there would be a DF epidemic exceeding the peak in 2019.
Conclusions: The study discloses a temporal periodicity inherent to the DF epidemic in GD. The three models are applicable for forecasting and evaluating the RP and RL of DF epidemic in GD, separately.

1. Introduction

Dengue fever (DF) is a viral disease, mainly transmitted by *Aedes aegypti* and *Aedes albopictus*, and is the most serious mosquito-borne infectious disease in the world. In the past 10 years, the incidence has shown a significant upward trend, and its prevention and control have become increasingly urgent, posing a major challenge to global public health. In 2019, DF was listed as one of the "top ten global health threats" by the World Health Organization [1]. In China, the DF epidemic has also been a major public health problem, and Guangdong (GD) province has faced the highest risk [2]. *Aedes aegypti* has become widespread in Yunnan province but

has diminished in Guangxi, Guangdong, and Hainan provinces in recent years. *Aedes albopictus* is distributed throughout mainland China, spanning 25 provinces and municipalities [3]. GD is located in the southernmost part of mainland China, along the southern coast, with a subtropical monsoon climate on the whole. It is suitable for breeding of *Aedes albopictus*. GD is a highly populated and developed province, with 21 prefecture-level cities, and 122 counties. It has a permanent population of over 120 million and has frequent exchanges with foreign countries. People are generally susceptible to dengue fever. Dengue fever was prevalent in GD mainly in Foshan City, in 1978, with 22,122 reported cases. Subsequently, the annual reported dengue fever cases exceeding 3000 in GD were 15,205 in 1980, 16,999 in 1981, 3335 in 1985, 5292 in 1986, 6812 in 1995, 45,188 in 2014, and 6101 in 2019, respectively [2,4,5]. During 2005–2020 in mainland China, there were 12,701 imported cases and 81,653 indigenous cases recorded. The indigenous cases were mainly clustered in GD (74.0 %). During this period, counties with high occurrence frequencies of indigenous dengue fever are mainly located in GD. Indigenous dengue fever occurred more than ten times in each of the seven counties of GD [3].

Worldwide DF research focuses on the DF virus evolution, epidemiology, immunology, pathogenesis, vector control, therapeutics, and vaccine research [6–8]. Some studies showed that the epidemiological characteristics of DF in GD have been undergoing new changes in the past 10 years, with a trend of multi-point outbreaks and an increasing trend of incidence. In 2014, a recorded historical maximum outbreak of DF occurred in GD with more than 45000 reported cases [5]. In the past 5 years up to 2021, the incidence of DF in 2019 was the highest. A few studies pointed out that the epidemic of DF in GD had a periodicity, and the outbreak interval was shortening [9,10]. However, this was only a simple and intuitive description based on the DF epidemic curve. The return period (RP) refers to the mean time before a random variable reaches or exceeds a certain threshold over a long period of time. The variable threshold is the return level (RL), also known as the return intensity [11]. The RP is commonly used to characterize the period of extreme weather events or extreme meteorological element threshold recurrence [12]. Nevertheless, there might be an important epidemiological feature of the RP and RL of DF epidemic in GD. An in-depth study on this issue to reveal the rules and characteristics of the RP and RL of DF may provide important reference information for GD health authorities to formulate strategies of DF prevention and control, as well as to prepare necessary health resources.

Diverse models have been employed for the forecast of dengue incidence, outbreaks, or risk assessment, including generalized linear regression models (GLRM) [13], seasonal autoregressive integrated moving average models (SARIMA), generalized additive models (GAM), artificial neural networks (ANN) [14], and spatiotemporal Bayesian models [15]. The above-mentioned models have been mainly used for short-term DF incidence or incidence rate prediction [14,15], and a few have been applied for short-term outbreak prediction [13,15]. But seldom no studies have dealt with the RP and the RL of DF epidemic. Chiu et al. [16] proposed quantile regression methods to determine the spatial distributions of DF incidence rates in various RPs, and RPs at distinct incidence rates in two areas in Taiwan. Lim et al. [11] used the Extreme Value Theory (EVT) model to study the RP of DF outbreak on provincial level in Thailand. Therefore, this paper aims to study the RP and RL of DF epidemic in GD.

In this study, three probability distribution models, namely lognormal distribution (Lognor D.), normal distribution (Norm D.), and generalized logistic distribution (GLD), were selected to fit the RP and RL of DF in GD. Norm D. and Lognor D. are the most commonly used models for studying the epidemiological characteristics of infectious diseases [11]. GLD is mostly used to describe the growth and development laws of biological populations and is also widely used in the medical field [17]. In recent years, they have been particularly useful in the research of COVID-19 epidemic trend [18,19]. Up to now, the reports on the above distribution models used for the study of the RP of infectious diseases are still rare. However, the DF epidemic has certain periodicity [9,10] and spatial-temporal heterogeneity [11]. Therefore, it is meaningful to use appropriate models to study the characteristics of RP and RL of the DF epidemic in GD, China.

2. Materials and methods

2.1. Data collection

In China, dengue fever is a notifiable disease, and all cases of dengue fever were diagnosed according to the unified diagnostic criteria issued by the Chinese Ministry of Health. Three editions of guidelines for dengue prevention and control were issued formally by the GD provincial public health authorities. The first version was released in 1979 and then updated in 1987 and 2015, respectively [20]. The reported DF cases in GD from 1978 to 2021 were collected through the Collection of Epidemic Data in Guangdong Province, the National Infectious Disease Reporting Information Management System accessed from Guangdong Provincial Center for Disease Control and Prevention, and also from some relevant publicly published papers [4,5,21,22]. The imported DF cases were excluded, and the indigenous cases were analyzed in this paper. We have obtained permission from the Guangdong Provincial Center for Disease Control and Prevention to use this data.

2.2. Model applicability evaluation methods

Regarding the definition of DF outbreak [23], in this study, the occurrence of 3 or more indigenous cases per year was defined as the occurrence of an epidemic. At first, the years that presented an annual number of indigenous cases of 3 or more were selected for analysis in the model directly. And then the sample selection rate was calculated (that is the sample size N selected for analysis divided by the total sample size N_0) for correction of analysis results. To meet the requirements of the model fitting conditions of independent identical distribution, a logarithmic transformation was applied to the data. Subsequently, the model fitting was conducted, and the results were obtained through the inverse logarithm transformation. Anderson Darling method (AD value) was used for data normality test. The applicability of the model was evaluated by drawing a probability diagram (P–P diagram). And we plotted the probability

density distribution curve of the annual case numbers based on the three models to further investigate the applicability of the model.

2.3. Probability density functions and guaranteeing probability functions of the three distribution models [24]

In this study, 13 probability distribution models, namely Norm D., Gumbel'l D., Exponweib D., Geneextreme D., Gumbel'r D., Weibull_max D., Weibull_min D., Logistic D., Pearson3 D., Gompertz D., Expon D., Lognorm D., and Genlogistic D. were used for a trial fitting separately, and finally the best 3 models, namely Norm D., Lognor D., and GLD were selected for analysis according to the goodness of fit with R^2, RMSE, and AIC.

The probability density function f(x) and the guaranteeing probability (GP) function $P(x \geq x_p)$ of Norm D. , Lognor D., and GLD, are shown in formula (1) (2), (3) (4), and (5) (6) respectively. Where x_p is x with a given probability.

$$f(x)_{NormD.} = \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}} \tag{1}$$

$$P\left(x \geq x_p\right)_{NormD.} = \int_{x_p}^{\infty} \frac{\exp(-x^2/2)}{\sqrt{2\pi}} dx \tag{2}$$

The probability density above is defined in the "standardized" form. To shift and/or scale the distribution use the loc and scale parameters. Specifically, norm.pdf(x, loc, scale) is identically equivalent to norm.pdf(y)/scale with y = (x - loc)/scale. The following four probability density functions were also defined in the "standardization" form.

$$f(x,s)_{LognorD.} = \frac{1}{sx\sqrt{2\pi}} \exp\left(-\frac{\log^2(x)}{2s^2}\right) \tag{3}$$

$$P\left(x \geq x_p\right)_{LognorD.} = \int_{x_p}^{\infty} \frac{1}{sx\sqrt{2\pi}} \exp\left(-\frac{\log^2(x)}{2s^2}\right) dx \tag{4}$$

Where x > 0, s > 0 and s is the shape parameter.

$$f(x,c)_{GLD} = c\frac{\exp\left(-x\right)}{(1+\exp\left(-x\right))^{c+1}} \tag{5}$$

$$P\left(x \geq x_p\right)_{GLD} = \int_{x_p}^{\infty} c\frac{\exp\left(-x\right)}{(1+\exp\left(-x\right))^{c+1}} dx \tag{6}$$

Where x ≥ 0, c > 0 and c is the shape parameter.

2.4. Evaluation methods of model goodness of fit

The coefficient of determination (R^2; (7)), the root mean squared error (RMSE; (8)), and the Akaike information criterion (AIC; (9)) were used to evaluate the goodness of fit of the models. Their calculation formulas are as follows [18,19,25]:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \widehat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \overline{y}_t)^2} \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\widehat{y}_t - y_t)^2} \tag{8}$$

$$AIC = 2k + n\,ln\left[\frac{\sum_{t=1}^{n}(\widehat{y}_t - y_t)^2}{n}\right] \tag{9}$$

Where, n is the sample size, k is the number of parameters, y_t is the observed value at t, $\widehat{y}_t$ is the fitted value at t, and $\overline{y}_t$ is the average of the observations.

The R^2 explains the variance of the model relative to the total variance. The value of R^2 is between 0 and 1. The higher the value, the better the goodness of fit. The value of R^2 is closer to 1, the model is more acceptable. The RMSE is the average of the squared errors of the residuals of a model. It can be used to compare the goodness of fit among different models. The lower value of RMSE indicates a better model. AIC measures the relative distance between the true likelihood function of the original data series and the built-in likelihood function of the model. A model with the minimum AIC value will be the best model [19,25].

2.5. Calculation of RP and RL

Assuming that the logarithm values of the numbers of annual DF cases conform to the normal probability density distribution, and the N (the number of measured years) values are sorted from small to large, then the empirical cumulative probability P_i ((10)) of the *i*th value is calculated as the follows [26]:

$$P_i = i/(N+1) \tag{10}$$

Where i is the order of the logarithm of the number of dengue fever cases in a year among N years. The quantile function is the inverse of the cumulative probability density distribution function. Thus, the corresponding quantile value will be calculated according to the given empirical cumulative probability.

Assuming that the annual number of cases x is a continuous random variable, and its probability density function is f(x). For any real number x, the cumulative probability distribution function less than x is represented by F (x) ((11)), that is [27,28].

$$F(x) = F(x < x_p) = \int_{-\infty}^{x_p} f(x)dx \tag{11}$$

Therefore, according to the empirical cumulative probability P_i of the ith observation value and the fitted distribution function F (x), the fitted value can be calculated corresponding to the ith observation value.

The probability of x_p occurrence with a RP ((12)) is the GP [$P(x \geq x_p)$], also known as exceeding probability (EP; (13))). The calculation formula is [27,28]

$$RP = 1 \ / \ P(x \geq x_p) \tag{12}$$

$$EP = P(x \geq x_p) = 1 - P(x < x_p) = 1 - F(x) = 1 \ / \ RP \tag{13}$$

Where, x_p is the RL of the RP. The EP corresponding to the RL of the RP of 50, 30, and 10 years is 1/50, 1/30, and 1/10, respectively.

In this study, because the case numbers in some years were less than 3, they were not directly included in the model fitting, so it was necessary to use the sample selection rate (N/N_0) to correct EP ((13)), that was,

$$EP_c = P(x \geq x_p)_c = 1 \ / \ (RP \times N \ / \ N_0) \tag{14}$$

And then EP_c ((14))were used to calculate the RP and RL.

2.6. Application of the fitting results

The RPs, RLs, and their 95 % CIs estimated with the three models were used to evaluate the RP of DF epidemic of different levels. Therefore, the RPs were predicted for the recorded historical DF epidemic maximum that happened in GD in 2014, and for the highest DF epidemic level that occurred in 2019 in the past 5 years up to 2021.

2.7. Data processing and visualization tools

In this study, Python language mainly with the packages of "NumPy", "Pandas", "Pylab", "Seaborn", "Matplotlib.pyplot", and "Scipy.stats" were used for data analysis and figure plotting [24,29]. All statistical analyses were two-sided and values of P < 0.05 were considered statistically significant.

Table 1
The number of indigenous dengue fever cases reported in Guangdong from 1978 to 2021.

Year	Number of cases	Year	Number of cases	Year	Number of cases	Year	Number of cases
1978	22122	**1989**	**0**	2000	384	2011	49
1979	635	1990	374	2001	342	2012	441
1980	15205	1991	371	2002	1342	2013	2811
1981	16999	**1992**	**2**	2003	37	**2014**	**45130**
1982	**0**	1993	359	2004	42	2015	1529
1983	**0**	1994	4	**2005**	**0**	2016	418
1984	**0**	1995	6812	2006	1008	2017	1470
1985	3335	**1996**	**0**	2007	377	2018	2990
1986	5292	1997	632	2008	66	**2019**	**4884**
1987	2601	1998	480	2009	7	2020	11
1988	**1**	1999	290	2010	111	**2021**	**1**

3. Results

3.1. Data assessment

The reported indigenous dengue cases in 44 years from 1978 to 2021 years was shown in Table 1. Among the 44 (N_0) years, 35(N) years had an annual number of DF cases higher than 2. Thus, 35 years of data were directly included in the model fitting, and the sample selection rate was $N/N_0 = 35/44 = 0.7955$. The sample composed of 35 numbers was tested for normality, AD = 6.65, P < 0.01. It indicated that the data did not conform to the normal distribution. Therefore, we performed the logarithmic transformation and retested the transformed data, which met the normal distribution with AD = 0.375 and P > 0.15.
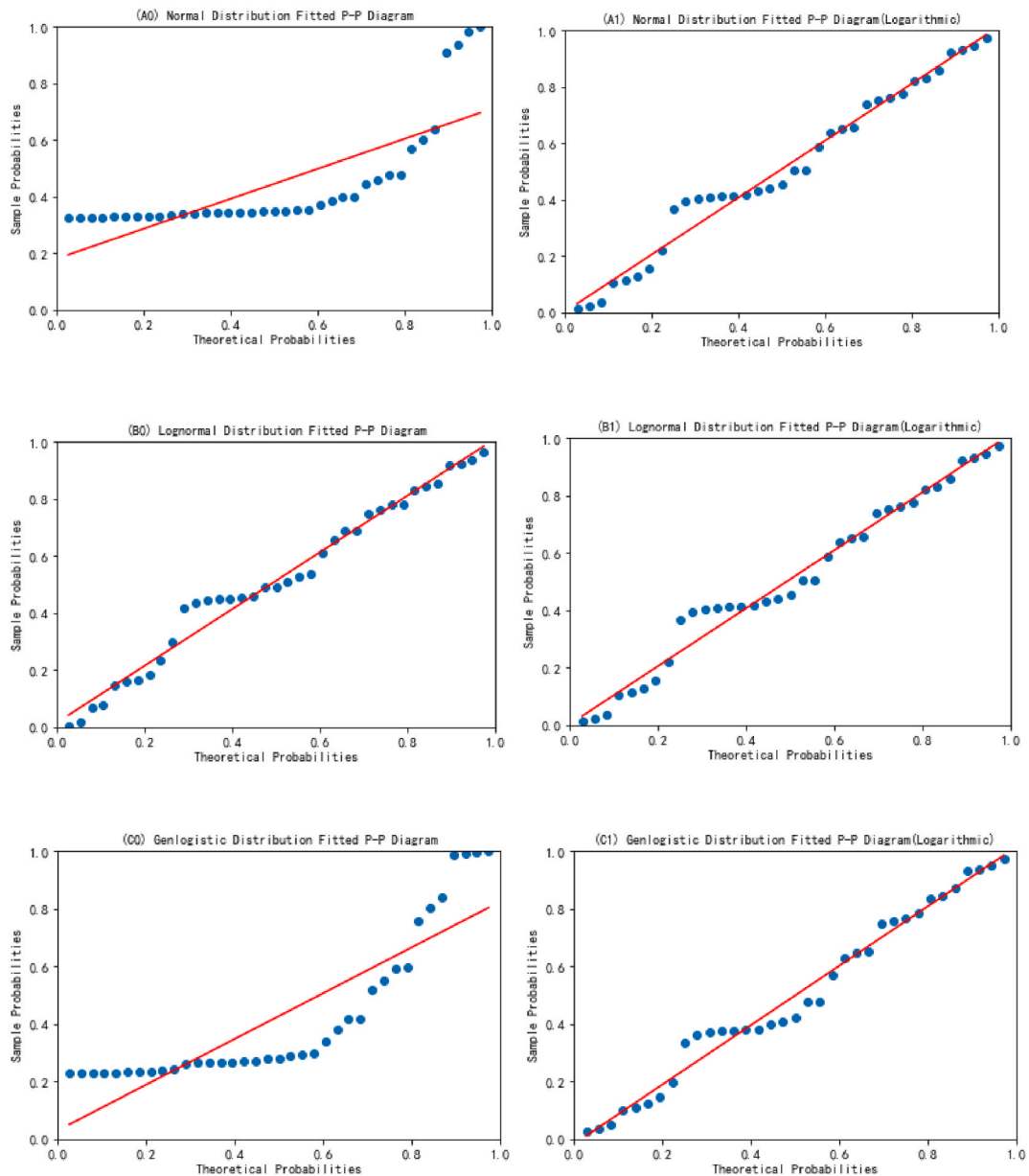

Fig. 1. The P–P diagram of the series data of annual case number fitted with the three distribution models respectively (A0, B0, and C0) and the P–P diagram of the logarithmic series data of annual case number fitted with the three distribution models respectively (A1, B1, and C1). Color should be used for Fig. 1 in print.

3.2. Model fitting suitability assessment results

Fig. 1 shows the one-to-one matched P–P diagrams of the series data of annual case number of DF sorted from small to large and that of the logarithmic series data, fitted with the three models respectively. The results showed that after logarithmic transformation of the annual number of cases, the P–P diagram fitted with the three models presented an approximate straight line correspondingly, indicating that the model fitting was applicable. And Fig. 2 shows the histogram and the probability density distribution curves of the logarithmic series data fitted with the three models. The results showed that the curves fitted by the three models were all basically "bell-shaped" after logarithmic transformation of the annual number of cases, which further indicated that the model fitting was applicable.

3.3. Evaluation of goodness of fit

Table 2 shows the goodness of fit of the three models. The goodness indicated that all three models fitted pretty well. Fig. 3 shows the comparison of the cumulative probability distribution curves of the sample values and the three model fitted values, individually. It shows that the goodness of fit of the model shown in Fig. 3 and Table 2 are highly consistent. Therefore, the models can be applied to predict and evaluate the RP of DF epidemic in GD.

3.4. Results of RP and RL

Table 3 show the RLs and their 95 % confidence intervals (CIs) of the case numbers of DF for different RPs of 2–45 years in GD fitted with the three models. According to the 95 % CIs, the RPs of the DF outbreak as large as what occurred in GD in 2014 (45130 cases) were 43 years (41365, 95 % CI: 36426–46973), 43years (41363, 95 % CI: 36425–46971), and 44 years (40809,95 % CI:36392–45762) assessed by the three models, separately. The RPs of the DF epidemic with the same prevalence intensity as that occurred in 2019 (4884 cases) were 7 years (4732, 95 % CI:4167–5374), 7 years (4732, 95 % CI: 4167–5374), and 8 years (5163, 95 % CI:4604–5790). Considering that 8 or 3 years have passed from 2014 or 2019 to 2022 and according to the result, we predicted by the three models independently, GD would have a large outbreak of DF that would exceed the historical maximum epidemic in the next 35, 35, or 36 years from 2022, and also in the next 4, 4, or 5 years from 2022, there would be a DF epidemic that would exceed the epidemic level occurred in 2019.

4. Discussion

This paper screened out 3 models with a criterion of R^2 above 0.95 from the 13 models. The assessment results showed that the annual DF case number sequence after logarithmic transformation essentially met the fitting conditions of independent and identical distribution. Logarithmic transformation of the data can significantly improve the applicability of fitting the three models. Therefore, the fitting analysis using the three models was appropriate.

As the main result of this study, a two-dimensional data table of the RPs, RLs,and their 95 % CIs was formed as a DF epidemic evaluation tool. The results showed that there were significant differences between the RLs of the two adjacent RPs with different time intervals (2–15 years apart) (95 % CI did not overlap), and it indicated that the DF epidemic in GD had a time periodicity of different epidemic intensity, so it told us there was a RP. The estimates of RLs, RPs, and EP are helpful for health resource planning purposes, especially for policy-makers [11]. Table 3 can be used as a professional tool to dynamically assess the intensity and transmission risk of DF outbreaks, assess the historical status of the epidemic, and evaluate the effectiveness of epidemic prevention and control measures. Therefore, it has important implications to study the RP and RL of DF epidemic in GD dynamically.

This study indicates that these models can be used to quickly assess how many years once of the epidemic intensity of DF had
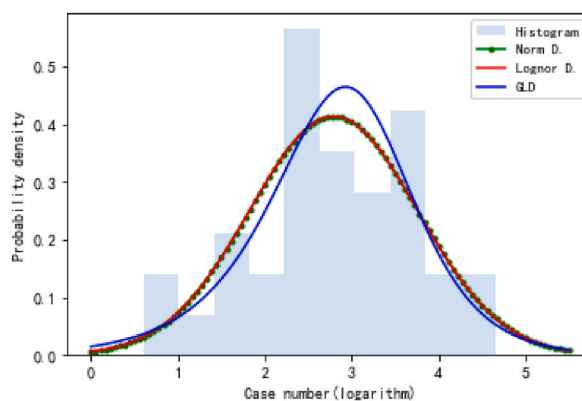

Fig. 2. Histogram of the annual case number (logarithm) of dengue fever and the probability density curves fitted by the three distributions, respectively. Color should be used for Fig. 2 in print.

Table 2
Parameter estimates and goodness of fit of the three models.

Models	Model parameter estimates (Logarithmic fitted)			Goodness of fit		
	location	scale	shape	R^2	RMSE	AIC
Lognor D.	1.21e-05	−7.98e04	7.98e04	**0.98**	1234.28	504.28
Norm D.	2.79	0.97		**0.98**	1234.40	502.28
GLD	0.71	3.09	0.48	**0.96**	1703.63	526.84

Lognor D., lognormal distribution; Norm D.,normal distribution; GLD, generalized logistic distribution; R^2, the coefficient of determination; RMSE, the root mean squared error; AIC, the Akaike information criterion.
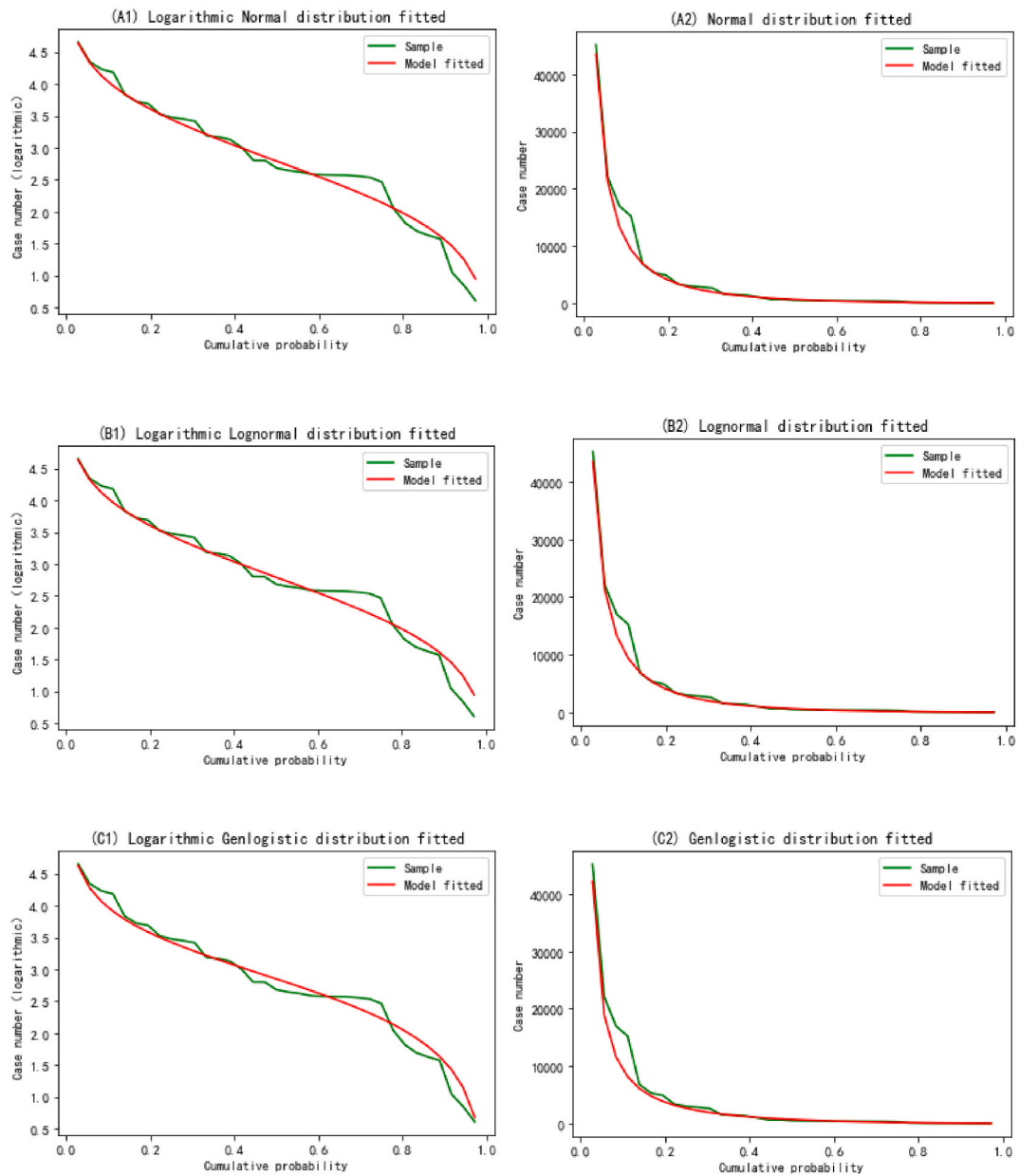

Fig. 3. Cumulative probability distribution curves of the sample values and the model fitted values. Color should be used for Fig. 3 in print.

Table 3
The RLs and their 95 % CIs of the DF case number for different RPs in GD fitted with the three models.

RPs	EPs (GPs)	EPcs	RLs(95 % CIs)		
			Lognor D.	Norm D.	GLD
2	0.500	0.629	297(262–337)	297(262–337)	362(322–405)
3	0.333	0.419	971(855-1102)	971(855-1102)	1056(942-1184)
4	0.250	0.314	1806(1591–2051)	1806(1591–2051)	1820(1623–2041)
5	0.200	0.251	2733(2407–3104)	2733(2407–3104)	2621(2338–2940)
6	0.167	0.210	3715(3272–4219)	3715(3272–4219)	3450(3076–3868)
7	**0.143**	**0.180**	**4732(4167–5374)**	**4732(4167–5374)**	4298(3833–4820)
8	**0.125**	**0.157**	5772(5083–6555)	5772(5083–6555)	**5163(4604–5790)**
9	0.111	0.140	6827(6012–7752)	6826(6011–7752)	6042(5389–6776)
10	0.100	0.126	7890(6948–8960)	7890(6948–8960)	6934(6184–7776)
11	0.091	0.114	8959(7890–10,174)	8959(7890–10,174)	7837(6989–8788)
12	0.083	0.105	10,031(8834–11,391)	10,031(8833–11,391)	8749(7803–9811)
13	0.077	0.097	11,104(9778–12,609)	11,104(9778–12,609)	9671(8624–10,845)
14	0.071	0.090	12,175(10,722–13,826)	12,175(10,722–13,826)	10,601(9454–11,888)
15	0.067	0.084	13,245(11,664–15,041)	13,245(11,664–15,041)	11,539(10,290–12,939)
16	0.062	0.079	14,312(12,604–16,253)	14,312(12,603–16,252)	12,483(11,132–13,998)
17	0.059	0.074	15,376(13,540–17,461)	15,376(13,540–17,460)	13,435(11,981–15,065)
18	0.056	0.070	16,436(14,473–18,664)	16,435(14,473–18,664)	14,392(12,835–16,139)
19	0.053	0.066	17,491(15,403–19,863)	17,491(15,403–19,862)	15,356(13,694–17,220)
20	0.050	0.063	18,542(16,329–21,056)	18,542(16,328–21,056)	16,325(14,558–18,307)
21	0.048	0.060	19,589(17,250–22,245)	19,588(17,250–22,244)	17,300(15,428–19,400)
22	0.045	0.057	20,630(18,167–23,427)	20,630(18,167–23,427)	18,280(16,301–20,498)
23	0.043	0.055	21,666(19,080–24,605)	21,666(19,080–24,604)	19,264(17,179–21,602)
24	0.042	0.052	22,699(19,989–25,776)	22,698(19,988–25,775)	20,254(18,062–22,712)
25	0.040	0.050	23,725(20,892–26,942)	23,724(20,892–26,941)	21,247(18,948–23,826)
26	0.038	0.048	24,747(21,792–28,102)	24,746(21,791–28,101)	22,245(19,838–24,945)
27	0.037	0.047	25,763(22,687–29,256)	25,762(22,686–29,255)	23,248(20,732–26,069)
28	0.036	0.045	26,774(23,578–30,404)	26,773(23,577–30,403)	24,254(21,629–27,198)
29	0.034	0.043	27,780(24,464–31,547)	27,780(24,463–31,546)	25,264(22,530–28,330)
30	0.033	0.042	28,782(25,345–32,684)	28,781(25,345–32,683)	26,278(23,434–29,467)
31	0.032	0.041	29,778(26,223–33,815)	29,777(26,222–33,814)	27,295(24,341–30,608)
32	0.031	0.039	30,769(27,096–34,941)	30,768(27,095–34,940)	28,316(25,252–31,753)
33	0.030	0.038	31,756(27,964–36,061)	31,755(27,963–36,060)	29,341(26,165–32,902)
34	0.029	0.037	32,738(28,829–37,176)	32,736(28,828–37,175)	30,368(27,082–34,054)
35	0.029	0.036	33,714(29,689–38,285)	33,713(29,688–38,284)	31,399(28,001–35,210)
36	0.028	0.035	34,687(30,545–39,389)	34,685(30,544–39,388)	32,433(28,923–36,370)
37	0.027	0.034	35,654(31,397–40,488)	35,653(31,396–40,487)	33,470(29,848–37,533)
38	0.026	0.033	36,617(32,245–41,582)	36,616(32,244–41,580)	34,511(30,775–38,699)
39	0.026	0.032	37,576(33,089–42,670)	37,574(33,088–42,668)	35,553(31,705–39,868)
40	0.025	0.031	38,530(33,929–43,753)	38,528(33,928–43,752)	36,599(32,638–41,041)
41	0.024	0.031	39,479(34,766–44,832)	39,477(34,764–44,830)	37,648(33,573–42,217)
42	0.024	0.030	40,424(35,598–45,905)	40,423(35,596–45,903)	38,699(34,511–43,396)
43	0.023	0.029	**41,365(36,426–46,973)**	**41,363(36,425–46,971)**	39,753(35,450–44,577)
44	0.023	0.029	42,302(37,251–48,037)	42,300(37,250–48,035)	**40,809(36,392–45,762)**
45	0.022	0.028	43,234(38,073–49,096)	43,233(38,071–49,094)	41,868(37,337–46,950)

RLs, the return levels; CIs, confidence intervals; DF, dengue fever; RPs, the return periods; GD, Guangdong province; EPs, exceeding probabilities; GPs, guaranteeing probabilities; EPcs, corrected EPs; Lognor D., lognormal distribution; Norm D., normal distribution; GLD, generalized logistic distribution.

occurred, and also be used to predict how many years an epidemic of certain intensity may recur in the future. However, it is noteworthy that the selection of fitting model for calculating the RP of a widespread outbreak of DF and the application of the fitted results should be more careful to prevent "overestimation" or "underestimation", to make more scientific evaluation and prediction.

Up to now, imported cases and local transmission caused by them are still typical features of DF epidemic in GD [5,30]. The number of analyzed cases in this study does not include imported cases. However, it is worth noting that GD has the natural conditions and vector for local transmission of DF [30]. The spread of DF is affected by mosquito vectors, dengue virus, climate, social environment and other factors [6,31]. With the warming of climate, the development of economy, trade and tourism, and the acceleration of urbanization, the global incidence rate of DF has increased significantly in recent decades [31]. Judging from the epidemic trend of DF in the world and in GD of China in the past 10 years, more attention should be paid to the possibility of an "extremely" high prevalence of DF epidemic in the future. Therefore, in the future, the evaluation and prediction of RP and RL of DF must take into account the above factors and the extrapolation ability of the model.

A key issue in public health resource planning is to predict the possibility of abnormal or extreme events in the future. The practice of dealing with the COVID-19 epidemic shows that we should use "probabilistic thinking" to deal with public health emergencies. Therefore, the study on the epidemic characteristics of infectious diseases, especially on the RP and RL of the outbreak, is greatly significant for the health departments to do a good job in the planning of prevention and control policy and epidemic risk management

and emergency response to the outbreak. This study is the beginning of the research on the RP of infectious diseases grounded in the theory of probability distribution. Up to now, no comparable research papers have been published in China, underscoring the urgent need to bolster such study.

The methods used in this study mainly include data screening, probability distribution test, logarithmic transformation of data, goodness of fit analysis and model screening, optimal probability model fitting and fitting result analysis. These methods have universal applicability for analyzing the return period of other similar diseases.

However, there are several limitations of this study. First, this study takes the provincial DF epidemic in GD as the study population, but there are some differences of occurrence frequencies of indigenous dengue fever among 21 cities, and among 122 counties in GD [2–5]. So this study can reveal the overall situation at the provincial level, but it cannot reveal its regional specificity. Second, this study is based on a data-driven model, which does not include factors affecting the epidemic situation in the model.

In conclusion, we demonstrate that the DF epidemic in GD has a temporal periodicity in this paper. The three models are applicable to predict and evaluate the RPs and RLs of DF epidemic in GD. The results show that GD is predicted to experience a DF outbreak beyond the historical maximum in the next 35 or 36 years from 2022. The findings offer valuable advisory insights to health authorities for strategic health resource allocation and preparedness in the prevention and control of DF, and may also provide a framework that enables health professionals to investigate the unique epidemiological features of relevant infectious diseases.

Funding

Ethical approval

Not relevant here for the current paper.

Data availability statement

All data is presented in Table 1 in text.

CRediT authorship contribution statement

Siqing Zeng: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Jianpeng Xiao:** Writing – review & editing, Validation, Data curation. **Fen Yang:** Writing – review & editing, Validation, Data curation. **Jiya Dai:** Writing – review & editing, Validation, Data curation. **Meng Zhang:** Writing – review & editing, Validation, Data curation. **Haojie Zhong:** Writing – review & editing, Validation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] World Health Organization. Ten threats to global health in 2019 [EB/OL].(2019-03-21) [2022-08-24].https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019.

[2] Qiyong Liu, Dengue fever in China: new epidemical trend, challenges and strategies for prevention and control, Chin. J. Vector Biol. Control 31 (1) (2020) 1–6, https://doi.org/10.11853/j.issn.1003.8280.2020.01.001.

[3] Yujuan Yue, Qiyong Liu, Xiaobo Liu, et al., Dengue fever in mainland China, 2005–2020: a descriptive analysis of dengue cases and Aedes data, Int. J. Environ. Res. Publ. Health 19 (2022) 3910, https://doi.org/10.3390/ijerph19073910.

[4] Zefeng Zeng, Jianfeng He, Analysis of dengue fever epidemic situation in Guangdong province from 1978 to 1999, Sci of Travel Med 6 (2) (2000) 1–5, https://doi.org/10.3969/j.issn.1006-7159.2000.02.001.

[5] Meng Zhang, Deng Aiping, Li Jiansen, et al., Epidemiological characteristics and trends of Dengue fever from 2012 to 2017 in Guangdong province of China, Chin J Viral Dis 8 (4) (2018) 282–287, https://doi.org/10.16505/j.2095-0136.2018.0067.

[6] J. Lourenço, W. Tennant, N.R. Faria, et al., Challenges in dengue research: a computational perspective, Evol Appl 11 (2018) 516–533, https://doi.org/10.1111/eva.12554.

[7] J.G. Low, E.E. Ooi, S.G. Vasudevan, Current status of dengue therapeutics research and development, J. Infect. Dis. 215 (suppl_2) (2017) s96–s102, 101093/infdis/jiw423.

[8] L.C. Katzelnick, J. Coloma, E. Harris, Dengue: knowledge gaps, unmet needs and research priorities, Lancet Infect. Dis. 17 (3) (2017) e88–e100, 101016/S1473-3099(16)30473-X.

[9] Xi Jiacheng, Xiaomin Cheng, Huan Hu, et al., Temporal dynamic of dengue fever in Guangdong province from 1990 to 2018, J. Trop. Med. 20 (4) (2020) 460–464, https://doi.org/10.3969/j.issn.1672-3619.2020.04.009.

[10] Xuefeng Hu, Shuang Wu, Yuanyuan Wu, et al., Quantity changes and analysis of dengue fever cases by time series fitting in mainland China from 1997 to 2020, Chin J Front Health Quarant 45 (3) (2022) 216–219, https://doi.org/10.16408/j.1004-9770.2022.03.011.

[11] J.T. Lim, B.S.L. Dickens, A.R. Cook, Modelling the epidemic extremities of dengue transmissions in Thailand, Epidemics-Neth 33 (2020) 100402, https://doi.org/10.1016/j.epidem.2020.100402.

[12] Li Shanshan, Zhengquan Cheng, Gong Yueting, Spatiotemporal distribution characteristics of tropical cyclone precipitation in Guangdong province and estimation of recurrence intervals, Guangdong Meteor 42 (4) (2020) 1–5, https://doi.org/10.3969/j.issn.1007-6190.2020.04.001.
[13] A.L. Ramadona, L. Lazuardi, Y.L. Hii, et al., Prediction of dengue outbreaks based on disease surveillance and meteorological data, PLoS One 11 (3) (2016) e0152688, https://doi.org/10.1371/journal.pone.0152688.
[14] O.S. Baquero, L.M. Santana, F. Chiaravalloti-Neto, Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models, PLoS One 13 (4) (2018) e0195065, https://doi.org/10.1371/journal.pone.0195065.
[15] R. Lowe, C. Barcellos, C.A. Coelho, et al., Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts, Lancet Infect. Dis. 14 (7) (2014) 619–626, https://doi.org/10.1016/S1473-3099(14)70781-9.
[16] Chuanhung Chiu, Tzaihung Wen, Lungchang Chien, Hualung Yu, A probabilistic spatial dengue fever risk assessment by a threshold-based-quantile regression method, PLoS One 9 (10) (2014) e106334, https://doi.org/10.1371/journal.pone.0106334.
[17] Minren Zhu, Fitting Gompertz model and logistic model, Math practice & theory 32 (5) (2002) 705–709, https://doi.org/10.3969/j.issn.1000-0984.2002.05.001.
[18] F.J. Diaz Perez, D. Chinarro, R.P. Otin, et al., Comparison of growth patterns of COVID-19 cases through the ARIMA and Gompertz models. case studies: Austria, Switzerland, and Israel, Rambam Maimonides Med J 11 (3) (2020), https://doi.org/10.5041/RMMJ.10413, 1–13.e0022.
[19] A.M.C.H. Attanayake, S.S.N. Perera, S. Jayasinghe, Phenomenological modelling of COVID-19 epidemics in SriLanka, Italy, the United States, and Hebei province of China, Comput Math Method Med (2020) 6397063, https://doi.org/10.1155/2020/6397063.
[20] Fengfu Cui, Feiwu He, Xiaorong Huang, et al., Dengue and dengue virus in Guangdong, China, 1978–2017: epidemiology, seroprevalence, evolution, and policies, Front. Med. 9 (2022) 797674, https://doi.org/10.3389/fmed.2022.797674.
[21] Liang Wenjia, Jianfeng He, Huiming Luo, et al., Epidemiological analysis of dengue fever in Guangdong province, 2001-2006, South China J. Prev. Med. 33 (5) (2007) 4–7, https://doi.org/10.3969/j.issn.1671-5039.2007.05.002.
[22] Runing Guo, Jianfeng He, Wenjia Liang, et al., Epidemiological analysis of foreign imported cases of dengue fever in Guangdong province, 2005-2010, South China J. Prev. Med 37 (5) (2011) 19–22. CNKI:SUN:GDWF.0.2011-05-004.
[23] Di Mu, Jinchao Cui, Wenwu Yin, et al., Epidemiological characteristics of dengue fever outbreaks in China, 2015-2018, Chin. J. Epidemiol. 41 (5) (2020) 685–689, https://doi.org/10.3760/cma.j.cn112338-20190715-00523.
[24] The SciPy community. Statistical functions (scipy.stats) [EB/OL]. [2022-09-15]. https://docs.scipy.org/doc/scipy/reference/stats.html#.
[25] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, Peer J Comput Sci 7 (2021) e623, https://doi.org/10.7717/peerj-cs.623.
[26] 1.6.11.1. Maximum wind speed prediction at the Sprogø station [EB/OL]. [2022-08-15]. http://scipy-lectures.org/intro/summary-exercises/stats-interpolate.html.
[27] Lisheng Tang, Hua Wang, Weiqin Liu, Liu Ye, Application of extreme value distribution theory in the forecast of chilling return periods of Guangdong province, China, Chin. J. Appl. Ecol. 29 (8) (2018) 2667–2674, https://doi.org/10.13287/j.1001-9332.201808.019.
[28] Li Qian, Yu Haiyang, Li Ting, et al., Hazard assessment of typhoons affecting the Beijing-Tianjin-Hebei region based on Gumbel distribution, Prog. Geogr. 37 (7) (2018) 933–945, https://doi.org/10.18306/dlkxjz.2018.07.007.
[29] The Matplotlib development team. Matplotlib: Visualization with Python [EB/OL]. [2022-09-15]. https://matplotlib.org.
[30] Fengfu Cui, The Study of Characteristics of Dengue Molecular Epidemiology and Prevention and Control Strategies in Guangdong Province, Southern Medical University, Guangzhou, 2022. II.
[31] Wei Yong, Xueli Zheng, Advances in the study of the factors influencing dengue fever transmission, J Pathogen Bio 16 (10) (2021) 1231–1235, https://doi.org/10.13350/j.cjpb.211025.