AUCTORES
Globalize your Research

Clinical Medical Reviews and Reports

Hiroshi Furutani *

Open Access

Research Article

Linear Regression Analysis of the COVID-19 Outbreak in Japanese Prefectures Using the Gompertz Distribution Model

Hiroshi Furutani [1]*, Tomoyuki Hiroyasu [2]

[1]AI x Humanity Research Center, Doshisha University, Kyotanabe City, 6100394, Kyoto, Japan.

[2]Department of Biomedical Sciences and Informatics, Doshisha University, Kyotanabe City, 6100394, Kyoto, Japan.

***Corresponding Author:** Hiroshi Furutani, AI x Humanity Research Center, Doshisha University, Kyotanabe City, 6100394, Kyoto, Japan.

Received date: September 13, 2025; **Accepted date:** September 26, 2025; **Published date:** October 13, 2025

Citation: Hiroshi Furutani, Tomoyuki Hiroyasu, (2025), Linear Regression Analysis of the COVID-19 Outbreak in Japanese Prefectures Using the Gompertz Distribution Model, *Clinical Medical Reviews and Reports,* 7(8); **DOI:10.31579/2690-8794/293**

Abstract

This paper reports a mathematical study analyzing the daily number of COVID-19 infections during the sixth wave (January to March 2022) in 13 Japanese prefectures. Using the Gompertz distribution model, which is commonly employed in biological studies, we estimate the peak position and height of the infection time series data. We apply a linear regression approach for parameter estimation and data fitting, focusing on the relationship between the peak time and the regression time window. The results show that the Gompertz model effectively estimates the height and position of the peak based on the provided data for Japanese prefectures.

Kew Words: COVID-19; linear regression; Gompertz; outbreak; Japan

Introduction

Coronaviruses are a diverse family of RNA viruses that infect humans, other mammals and avian species, including livestock. Infections caused by human and animal coronaviruses primarily result in respiratory and enteric diseases. Human coronaviruses have long been known to circulate in the population, typically causes seasonal and generally mild infections. In contrast, SARS-CoV, MERS-CoV and SARS-CoV-2 are highly pathogenic [1, 2]. Consequently, research on coronavirus epidemics has become a critical area of study.

The logistic distribution has been a widely used and important model for describing disease progression because of its effectiveness in characterizing many epidemics [3]. Originally, Verhulst proposed it to model human population growth. Interestingly, more than a decade earlier, Gompertz introduced a model for ani- mal population growth [4]. Although the Gompertz model has been used far less frequently than the logistic model, several studies highlight its usefulness [5, 6].

Time series data on COVID-19 from many countries and regions often exhibit a right-skewed distribution. Similarly, the density function of the Gompertz distribution has a right-skewed form. Consequently, several studies have applied the Gompertz function to model COVID-19 data [7, 8]. In our research, we also utilized the Gompertz distribution to analyze COVID-19 infections and deaths [9, 10].

In Japan, the surge of COVID- 19 cases from January to March 2022 has been commonly referred to as the sixth wave of the outbreak. This paper investigates the daily number of cases reported in thirteen prefectures during this period. In a previous study [11], we applied a linear regression method to estimate model parameters and fit the reported death data from foreign countries. Similarly, this paper employs a linear regression model within the framework of the Gompertz distribution. Our analysis focuses on the peak properties of COVID-19 time series data. To examine the effects of the distance between the peak date and the regression time window, we perform regression analyses via two-time windows. Addition- ally, as a special case, we conduct calculations with three-time windows for Tochigi Prefecture.

We reported preliminary results at symposium [12].

Dataset

Japan is divided into 47 prefectures with a total population of approximately 126 million people. The three major metropolitan areas are centered around Tokyo, Osaka, and Nagoya. This study focuses on 13 prefectures; Tokyo, Kanagawa and Saitama from the Tokyo metropolitan area; Osaka, Hyogo and Wakayama from the Osaka metropolitan area; and Aichi, Gifu and Shizuoka from the Nagoya metropolitan area. Additionally, Hokkaido, Hiroshima, Fukuoka and Tochigi are selected from other regions. Hokkaido, Hiroshima and Fukuoka are among the prefectures with large populations. Hokkaido is included because it is geographically isolated from the other 46 prefectures. Tochigi is chosen because of the necessity of carefully examining the relation- ship between theoretical predictions and observed data in this region. Together, the 13 selected prefectures represent approximately 60% of Japan's total population.

We downloaded the database of the NHK website (2022); https://www3.nhk.or.jp/news/special/coronavirus/data/" nhk news covid19 prefectures daily data.csv", downloaded on April 4, 2022.

The database contains daily numbers of COVID-19 infections and deaths from January 2020 until April 2022. We use the daily number of infections from September 1, 2021, to April 3, 2022.

Method

The cumulative distribution function of the Gompertz distribution is defined as

$$F(t) = \exp\{-e^{-y(t)}\},\ \ y(t) = a(t-b). \quad (1)$$

The probability density function of the Gompertz distribution is given by

$$f(t) = ae^{-y(t)}F(t). \quad (2)$$

Then we have

$$f(t)/F(t) = ae^{-y(t)}. \quad (3)$$

In the analysis, we use this relation.

Let *U (t)* be the cumulative number of infections on the t-th day. Since reported data of daily count *u(t)* often fluctuate significantly, we use the 7-day moving average *m(t)* as the daily count. We set t = 1 at the maximum value of *m(t).*

For Gompertz model analysis, three parameters must be estimated: the total number N, the shape parameter a, and the position parameter *b*. The parameter N can be eliminated via the relations

$$U(t) = NF(t), \qquad m(t) = Nf(t). \quad (4)$$

Our method uses the value $M(t)$, which is defined as

$$M(t) = m(t)/U(t). \quad (5)$$

From eq. (4), the value $M(t)$ can be estimated as

$$M(t) \approx \{Nf(t)\}/\{NF(t)\}. \quad (6)$$

Thus, we have

$$M(t) \approx ae^{-a(t-b)}. \quad (7)$$

This allows us to estimate $M(t)$ from the reported daily numbers.

The final task is to estimate three parameters. By applying a logarithmic transformation, we define $L(t)$ as

$$L(t) = -\ln M(t) \approx -\ln a + a\ (t-b). \quad (8)$$

Thus, $L(t)$ can be approximated by a linear function of t, allowing the parameters a and b to be estimated via the linear regression method. This study performs the regression analysis via a time window of 12 elements, denoted as W[ts,te=ts+11]. The next step is to estimate the total number

N. We use the average ratio over the interval [ts, ts+11]

$$N = \text{Ave}\ \{U(t)/F(t)\}. \quad (9)$$

Details of the estimation process can be found in a previous paper [10].

Results

Tokyo metropolitan area

First, we analyzed the daily infection numbers in Tokyo, Kanagawa, and Saitama during the sixth outbreak. These prefectures collectively account for 24.5% of Japan's population. Day 1 corresponds to February 5 for Tokyo, February 7 for Kanagawa, and February 8 for Saitama.

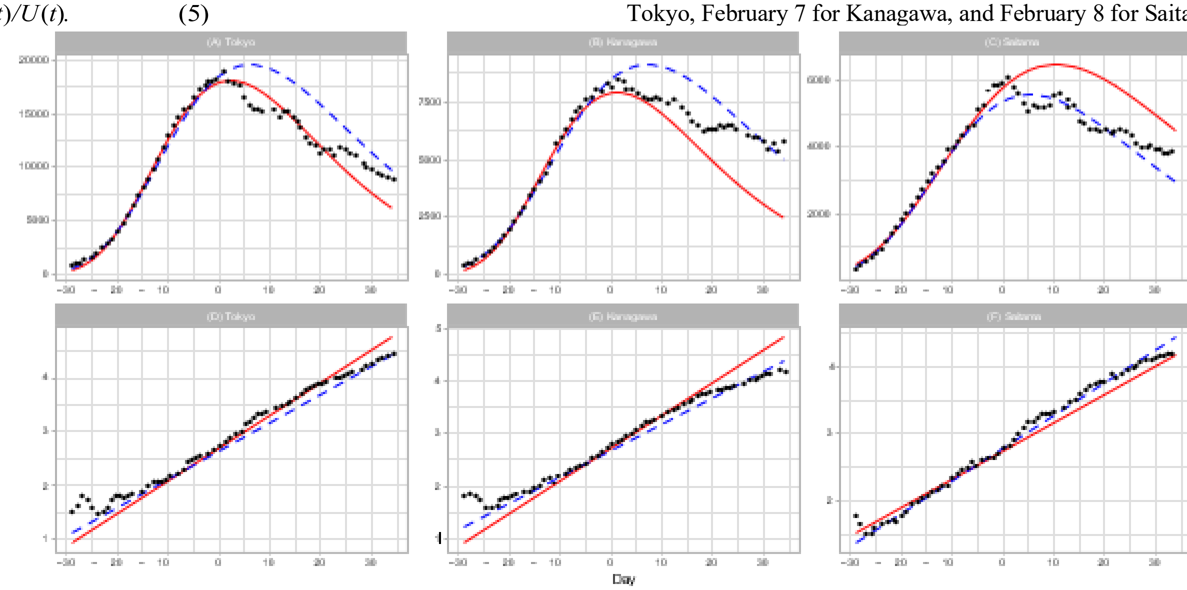

Figure 1: Results for Tokyo, Kanagawa, and Saitama. Upper panels, plots of the daily number of infections. Lower panels, plots of linear regression lines. The red solid line represents W [-11,0], and the blue dashed line represents W [-15, -4]. Closed circles indicate reported data.

Table 1 presents the estimated values for peak height and peak position using two-time windows for regression analysis: W [−11, 0] and W [−15, −4]. Note: The column labeled m represents the maximum number of reported cases, me is the estimated value of m, and b is the estimated peak position corresponding to t = 1

Prefecture	Window	*m*	*me*	*b*
Tokyo	W [-11,0]	18,565	18,027	1.75
	W [-15, -4]		19,535	5.54
Kanagawa	W [-11,0]	8,256	7,958	0.69
	W [-15, -4]		9,114	6.64
Saitama	W [-11,0]	6,053	6,478	9.88

	W [-15, -4]		5,615	4.84

Table 1: Results for Tokyo, Kanagawa, and Saitama.

Figure 1 illustrates the reported data and the Gompertz model estimation for Tokyo (left), Kanagawa (center), and Saitama (right). For Tokyo and Kanagawa, the Gompertz model curves using W [−11, 0] align well with the reported data, whereas the results using W [−15, −4] show some deviation. Conversely, for Saitama, the analysis using W [−15, −4] provides a better fit to the reported data than that using W [−11, 0].

Osaka metropolitan area

This subsection presents the results of Gompertz model analysis for Osaka, Hyogo, and Wakayama. These prefectures collectively account for 12.2% of Japan's population. Day 1 corresponds to February 8 for Osaka, February 7 for Hyogo, and February 5 for Wakayama.

Table 2 lists the numerical results of the regression analysis using W [−11, 0] and W [−15, −4].

Note: The column labeled m represents the maximum number of reported cases, me is the estimated value of m, and b is the estimated peak position corresponding to t = 1.

Prefecture	Window	*m*	*me*	*b*
Osaka	W [-11,0]	13,351	13,143	1.66
	W [-15, -4]		12,914	3.24
Hyogo	W [-11,0]	5,613	5,735	4.63
	W [-15, -4]		5,999	6.63
Wakayama	W [-11,0]	510	493	0.74
	W [-15, -4]		585	7.28

Table 2: Results for Osaka, Hyogo, and Wakayama.

Figure 2 illustrates the daily numbers of reported cases alongside the theoretical estimations for Osaka (left), Hyogo (center), and Wakayama (right). For Osaka, the theoretical estimation reproduces the reported data very well. Notably, the theoretical results obtained via W [−11, 0] and W [−15, −4] are almost identical in their estimates of the reported data. For Hyogo, the results of Gompertz model analysis using W [−11, 0] and W [−15, −4] demonstrate a reasonable fit to the data. Similar to Osaka, the two theoretical curves closely resemble each other. For Wakayama, the Gompertz model estimation using W [−11, 0] reproduces the daily number of infections very accurately. However, the theoretical curve generated with W [−15, −4] overestimates the reported data in the decreasing phase.

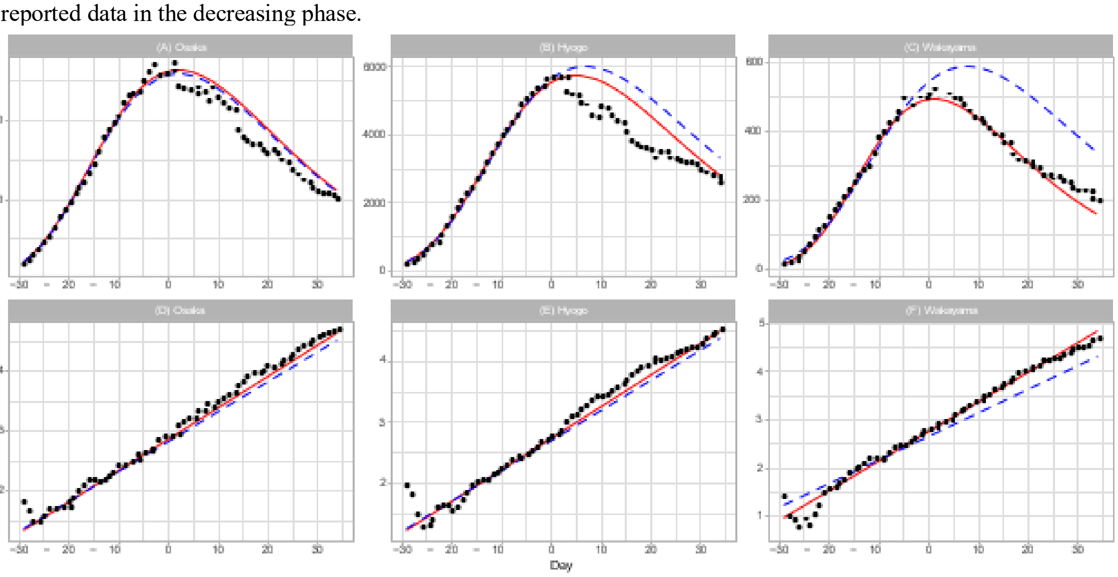

Figure 2: Results for Osaka, Hyogo, and Wakayama. The upper panels are plots of the daily number of infections. The lower panels are plots of linear regression lines. The red solid line represents W [-11,0], and the blue dashed line represents W [-15, -4]. Closed circles indicate reported data.

Nagoya metropolitan area

This subsection presents the results of the Gompertz model analysis for Aichi, Gifu, and Shizuoka. The population of these prefectures' accounts for 10.6% of Japan's total population. Day 1 corresponds to February 16 for Aichi, February 16 for Gifu, and February 7 for Shizuoka.

Table 3 lists the results of the Gompertz model analysis using W [−11, 0] and W [−15, −4].

Note: The column labeled m represents the maximum number of reported cases, me is the estimated value of m, and b is the estimated peak position corresponding to t = 1.

Prefecture	Window	*m*	*me*	*b*
Aichi	W [-11,0]	5,953	5,828	-2.90
	W [-15, -4]		5,707	-8.63
Gifu	W [-11,0]	943	905	-0.41
	W [-15, -4]		867	-9.15

Shizuoka	W [-11,0]	1,735	1,620	-0.12
	W [-15, -4]		1,495	-4.62

Table 3: Results for Aichi, Gifu, and Shizuoka.

Figure 3 shows the reported data and the Gompertz model estimation for Aichi (left), Gifu (center), and

Shizuoka (right). The analysis using W [−11, 0] provides good estimates for the peak height (m) and peak position (b) in all three prefectures. For Aichi, the peak at t = 1 has a bumpy structure. In the region of t ≤ −7, the analysis with W [−15, −4] offers a better fit to the reported data than W [−11, 0]. For Gifu, the infection curve exhibits two peaks; the first peak occurs at t = −7, and the second peak occurs at t = 1, showing a bumpy structure similar to Aichi. The analysis using W [−15, −4] appears to capture the first peak, whereas W [−11, 0] captures the second peak. For Shizuoka, Table 3 indicates that the analysis using W [−11, 0] provides reasonable estimates for the peak height (m) and the position of t = 1. However, the fitting quality of the analysis using W [-15, -4] is relatively poor, as the model underestimates the peak height.

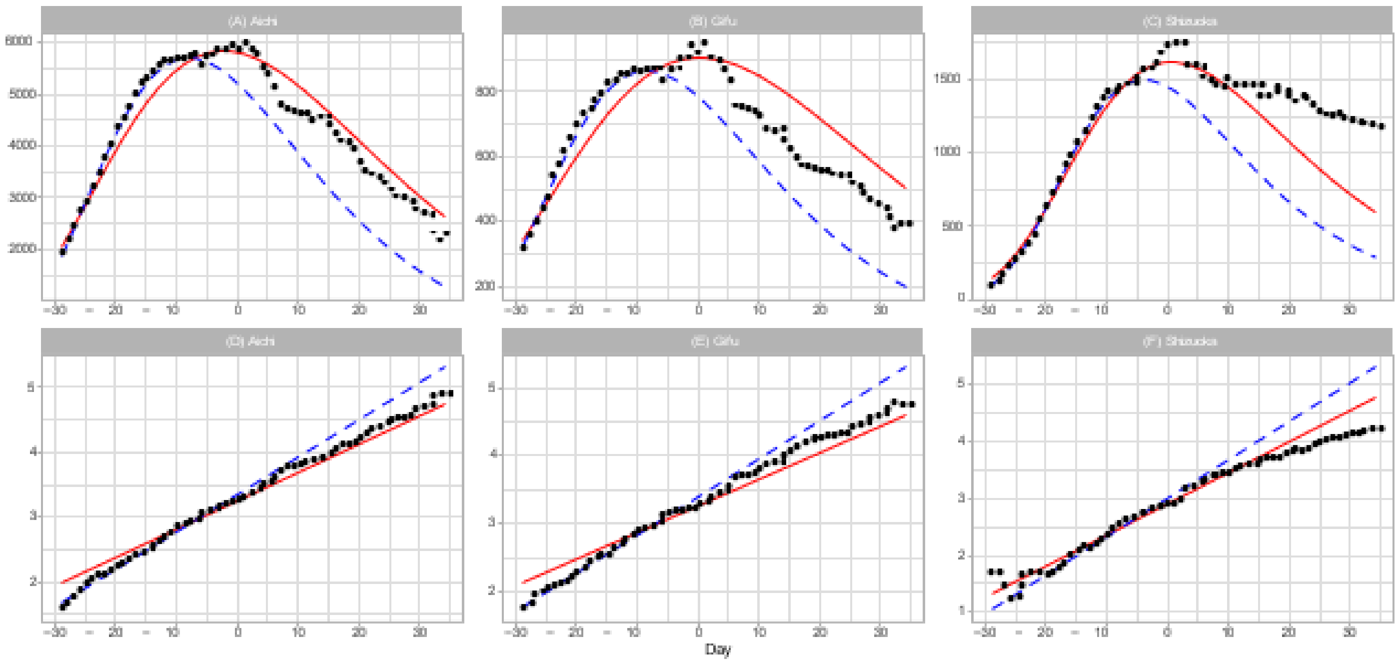

Figure 3: Results for Aichi, Gifu, and Shizuoka. The upper panels are plots of the daily number of infections. The lower panels are plots of linear regression lines. The red solid line represents W [-11,0], and the blue dashed line represents W [-15, -4]. Closed circles indicate reported data.

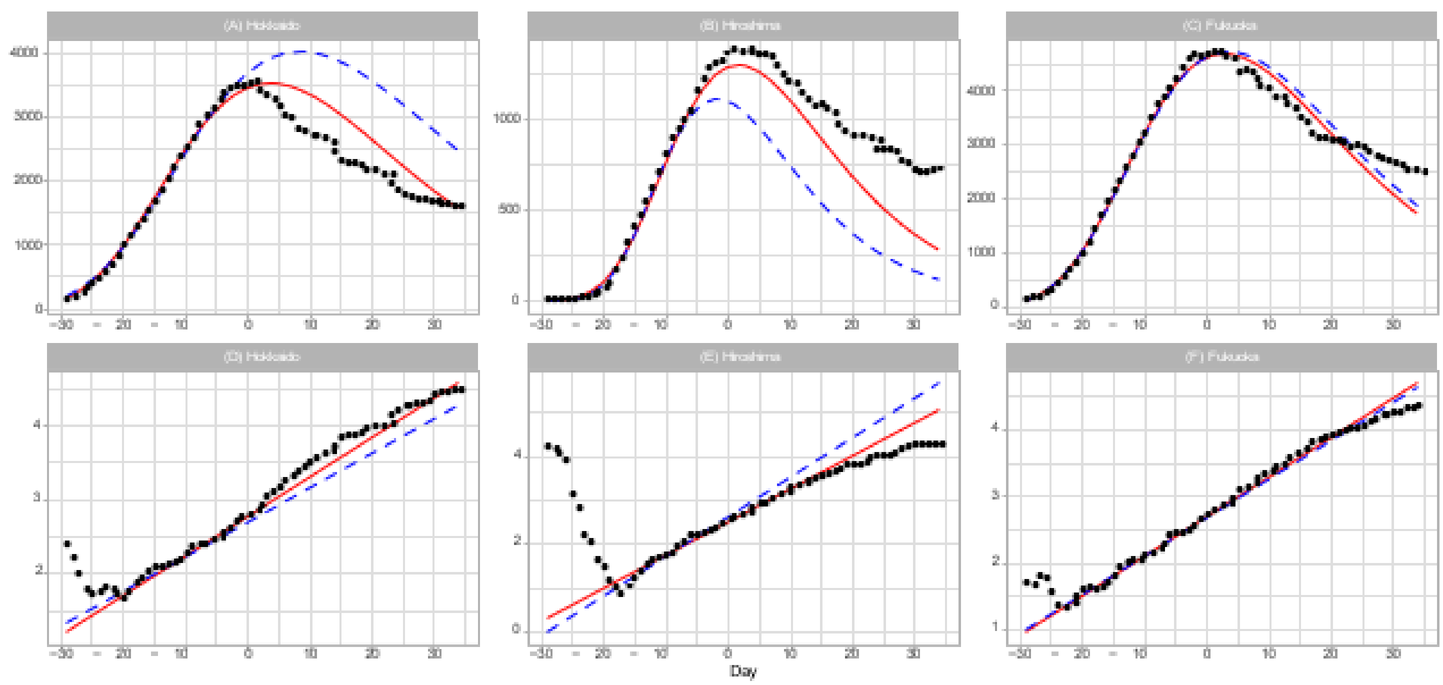

Figure 4: Results for Hokkaido, Hiroshima, and Fukuoka. The upper panels are plots of the daily number of infections. The lower panels are plots of linear regression lines. The red solid line represents W [-11,0], and the blue dashed line represents W [-15, -4]. Closed circles indicate reported data.

Other regions

This subsection presents the results of Gompertz model analysis for Hokkaido, Hiroshima, Fukuoka, and Tochigi. Hokkaido is located on Hokkaido Island, which is isolated from mainland Japan, whereas Fukuoka is located on Kyushu Island, the second largest island in Japan. Hiroshima and Tochigi are located on mainland Japan. The combined population of these four prefectures accounts for 12.0% of Japan's total population. Day 1 corresponds to February 8 for Hokkaido, January 23 for Hiroshima, February 5 for Fukuoka, and February 17 for Tochigi.

Table 4 presents the results of the Gompertz model analysis for Hokkaido, Hiroshima, and Fukuoka using W [−11, 0] and W [−15, −4].

Note: The column labeled m represents the maximum number of reported cases, me is the estimated value of m, and b is the estimated peak position corresponding to t = 1.

Prefecture	Window	*m*	*me*	*b*
Hokkaido	W [-11,0]	3,494	3,550	3.33
	W [-15, -4]		4,016	8.10
Hiroshima	W [-11,0]	1,366	1,301	1.24
	W [-15, -4]		1,106	-2.15
Fukuoka	W [-11,0]	4,708	4,700	2.42
	W [-15, -4]		4,716	3.12

Table 4: Results for Hokkaido, Hiroshima, and Fukuoka.

In Figure 4, graphs of the reported data and theoretical results are plot- ted for Hokkaido (left), Hiroshima (center), and Fukuoka (right). For Hokkaido and Hiroshima, the model calculations using W [−11, 0] closely reproduce the reported data. How- ever, the calculation with W [−15, −4] for Hokkaido overestimates the reported infection curve, whereas for Hiroshima, it underestimates the reported curve. For Fukuoka, the analysis using both W [−11, 0] and W [−15, −4] provides a reasonable fit to the reported curve.

Table 5 lists the Gompertz model estimates using W [−11, 0], W [−15, −4], and W [−19, −8] for Tochigi.

Note: The column labeled m represents the maximum number of reported cases, me is the estimated value of m, and b is the estimated peak position corresponding to t = 1.

Prefecture	Window	*m*	*me*	*b*
Tochigi	W [-11,0]	850	785	-2.80
	W [-15, -4]		796	-12.29
	W [-19, -8]		807	-6.92

Table 5: Results for Tochigi.

Figure 5 shows the daily reported cases alongside the theoretical results for Tochigi. A double-peaked epi-demic curve is observed. The first (lower) peak occurs at t = −11, with a peak height of 819. Table 5 shows that the Gompertz model estimate me using W [−19, −8] provides a bet- ter fit than those using W [−11, 0] and W [−15, −4]. The estimates b = −12 and me = 796 using W [−15, −4] seem to correspond to the first peak.

Discussion

In the period of the sixth wave outbreak, sub-emergency measures were issued for many prefectures. In Tokyo, sub-emergency measures were applied from January 21 till March 21, or −14 ≤ t ≤ 45. For Osaka, sub-emergency measures were issued in the period −11 ≤ t ≤ 27, and for Aichi, −25 ≤ t ≤ 34. For Hiroshima, sub-emergency measures were applied January 9 till March 6, or −15 ≤ t ≤ 41. We have checked the linearity of L(t) for 14 prefectures, and have investigated the relation- ship to the sub-emergency measures. However, we can find no systematic relationship between them.

Notably, our approach closely resembles the K-value model pro- posed by Nakano and Ikeda [7]. The K-value is an indicator of the spread rate and is defined as

$$K(t) = 1 - \frac{U(t-7)}{U(t)}. \qquad (10)$$

After some modifications

$$K(t) = \{U(t) - U(t-7)\}/U(t),$$

we have a relation using a seven-day moving average

$$K(t) = \frac{7\bar{m}\,(t-3)}{U(t)}.$$

Thus, we can show

$K(t) \approx 7M(t)$,

where M (t) is defined in eq. (5) of the Gompertz model.

The Gompertz function has been shown to fit the growth curve of chicken heart [13]. Here, we present his derivation of the Gompertz model. Let ϕ (U) be the growth rate defined

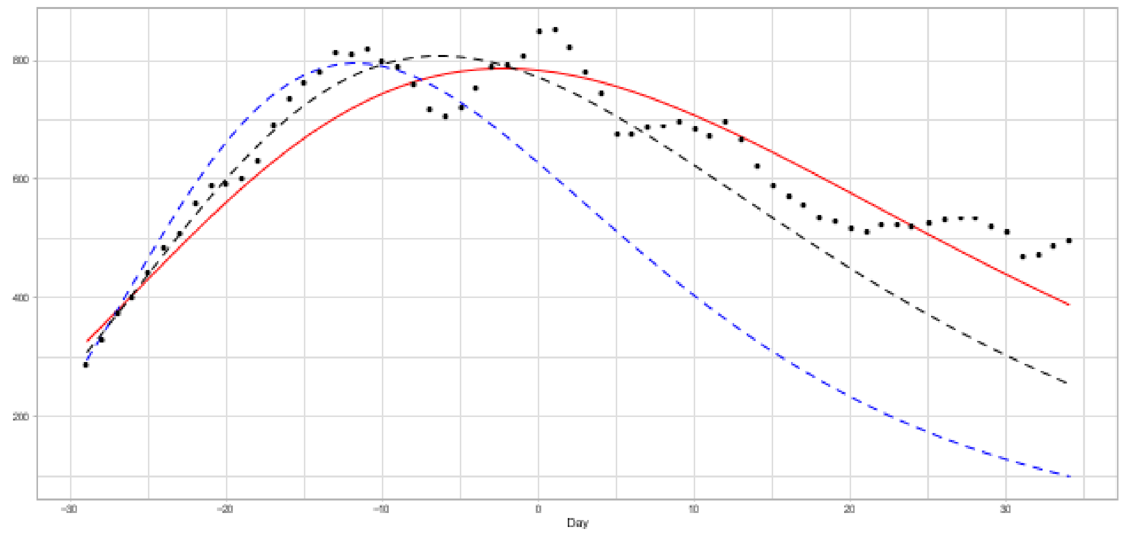

Figure 5: Daily numbers of Tochigi. The red solid line represents W [-11,0], the blue dashed line represents W [-15, -4], and the black dashed line represents W [-19, -8] Closed circles indicate reported data.

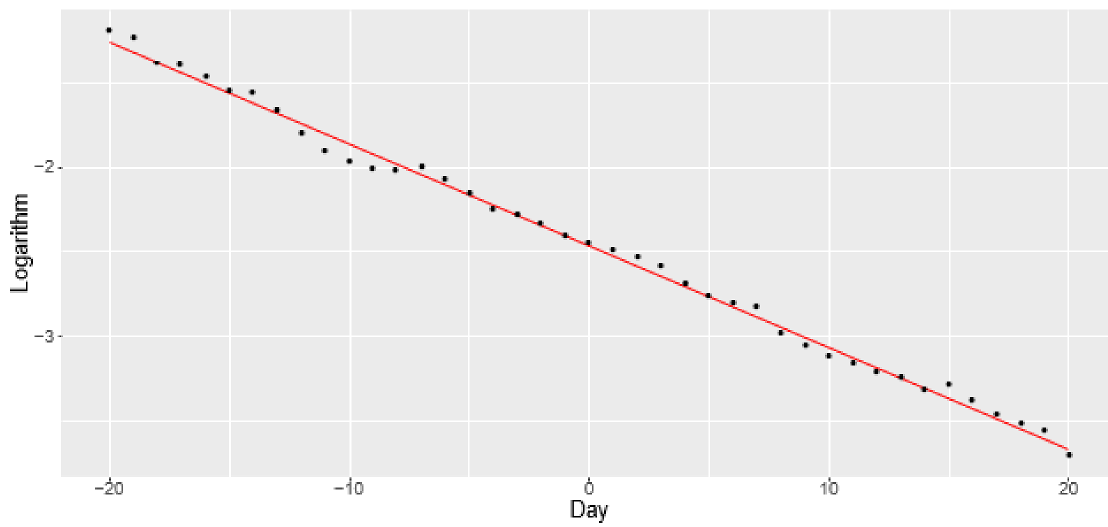

Figure 6: Logarithm of growth rate for Osaka

As

$$\varphi(U)= \frac{U(t+1)-U(t)}{U(t)}.$$

He has shown that ϕ can be approximately given by a decreasing exponential function of time

$$\phi\ (U)=Ae^{-at}, \qquad (11)$$

with constants A and a. Figure 6 shows the logarithm of the growth rate of infections in Osaka, where the daily count U (t +1) −U (t) is replaced by the moving average m (t + 1).

The growth rate ϕ may be approximated by

$$\varphi(U)\approx \frac{1}{U(t)}\frac{d}{dt}U(t)=\frac{d}{dt}\ln U(t).$$

We integrate the equation

$$\frac{d}{dt}\ln U(t) = Ae^{-at},$$

and obtain the solution with the form of the Gompertz function

$$U\ (t)\ =\ C\ \exp\{-Be^{-at}\}, \quad (12)$$

where B = A/a. The constant C depends on the initial condition.

We consider the application of a flexible growth function proposed by Richards [14]. Zou et al. reported successful results when a logistic growth model was used to describe the COVID-19 outbreak in China [15]. Richards' function can describe a wide range of distributions, including logistic and Gompertz distributions, as sub-models. The cumulative distribution functions of the logistic and Gompertz models can be transformed to the function z(t)

$$z(t) = \alpha + \beta\rho^t, \qquad (13)$$

where α, β and ρ are parameters. We are now analyzing the COVID- 19 outbreak data using a non-linear regression method.

Conclusion

We present the analysis of infection counts for 13 prefectures during the sixth wave of the COVID-19 outbreaks in Japan. This study encompasses approximately 60% of the Japanese population. Following previous research, we modeled the daily number of infections via the Gompertz distribution. Using the linear regression approach with the time window of W [-11,0], we accurately estimate the peak height and peak date of the reported infection data. However, the estimation quality for W [-15, -4] varies across prefectures. Notably, our analysis of Tochigi reveals that the time window W [-19, -8] provides better estimates of peak characteristics than W [-11,0]. This finding indicates that the estimation using the time window of the early phase must be carefully examined.

Overall, our analysis demonstrates that modeling epidemic curves within the framework of the Gompertz function holds great promise for studying COVID-19 dynamics in Japanese prefectures.

Data availability

The data generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Author contributions

The present study was conducted equally by the authors.

Competing interests

The authors declare no competing interests.

References

1. Masters, P.S. (2006). The molec- ular biology of coronaviruses. Adv Virus Res, 66,193-292.
2. V'kovski, P., Kratzel, A.,Steiner, S., Stadler, H. & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol, 19:155-170.
3. Madden, L.V. & Campbell, C.L. (1990). Nonlinear disease progress curves. In Klantz, J. (ed.) Epidemics of Plant Diseases, 181-229.
4. Gompertz, B. (1815). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Phil Trans R Soc Lond, 115:513-585.
5. Berger, R.D. (1981). Comparison of the Gompertz and logistic equations to describe dis- ease progress. Phytopathology, 71:716-719.
6. Fleming, R.A. (1983). Development of a simple mechanistic model of cereal rust progress. Phytopathology, 73, 308-312.
7. Nakano, T. & Ikeda, Y. (2020). Novel indicator to ascertain the status and trend of COVID-19 spread: Modeling study. J Med internet Res, 22:20144.
8. Ohnishi, A., Namekawa, Y. & Fukui, T. (2020). Universality in COVID-19 spread in view of the Gompertz function. Prog Theor Exp Phys, 123J10.
9. Furutani, H., Hiroyasu, T. & Okuhara, Y. (2022). Method for estimating time series data of COVID-19 deaths using a Gumbel model. Archi Clin Biomed Res, 6:50-64.
10. Furutani, H. & Hiroyasu, T. (2022). Estimation of COVID- 19 cases in Japanese prefectures using a Gumbel distribution. Archi Clin and Biomed Res, 6, 756-763.
11. Furutani, H. & Hiroyasu, T. (2023). Linear regression analysis of COVID-19 time-series data using the Gumbel distribution. J Infect Dis Ther, 11:1000553.
12. Furutani, H. & Hiroyasu, T. (2023). Analysis of the sixth wave COVID-19 outbreak in Japan. Proc ISAROB.
13. Medawar, P.B. (1940). The growth, growth energy, and aging of the chicken's heart. Proc R Soc Lond Ser B, 129, 332-355.
14. Richards, F.J. (1959). A flexible growth function for empirical use. J Exper Botany, 10:290-301.
15. Zou, P., Pan, S., Zhao, P., Han, L., Wang, X., et al. (2020). Outbreak analysis with a logistic growth model shows COVID-19 suppression dynamics in China. PLoS ONE, 15:0235247

DOI:10.31579/2690-8794/293