Probabilistic Analysis Depending on the Distance from A COVID-19 Outbreak

Yupaporn Areepong [1] , Rapin Sunthornwat [2*]

[1] *Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok,10800, Thailand.*

[2] *Industrial Technology and Innovation Management Program, Pathumwan Institute of Technology, Bangkok, 10330, Thailand.*

Abstract

COVID-19 has been affecting human beings since the end of 2019. Studying the characteristics of a COVID-19 outbreak is significant because it will add to the knowledge that is necessary for protecting the general public and controlling future viral outbreaks. The aims of the present research are to analyze COVID-19 outbreaks in Thailand depending on the distance from the outbreak center by using a differential equation, to construct a probability density function from the solution of the differential equation, and to prove the theorem for the probability density function depending on the distance from the outbreak. The least-squares-error method is adopted to estimate the parameters of the function describing the COVID-19 outbreak. Moreover, a cumulative distribution function, a quantile function, a sojourn function, a hazard function, the median, the expected value, variance, skewness, and kurtosis are derived, and their practicability is shown. Applying the exponentially weighted moving average control chart to monitor a COVID-19 outbreak based on distance is proposed and compared with monitoring the COVID-19 outbreak based on time. The results show that using the former more quickly detected the out-of-control first passage time of the COVID-19 outbreak than the latter.

1- Introduction

COVID-19, which first appeared in late 2019, has had a severe impact globally; it has caused health and economic problems all around the world, and national governments in every country have tackled the problems by developing policies and vaccines to fight it. Governments have to make complex decisions during any viral outbreak situation, and the COVID-19 outbreak was no exception. Decisions made by government officials need to be informed ones, and that requires accurate predictive models and data analysis. In particular, estimation and forecasting models are the key data analytic methods for describing and predicting the daily spread of a virus during an epidemic. If their predictability is precise, they can help save lives. Most forecasting and mathematical models for describing the spreading characteristics of a COVID-19 outbreak, such as the number of cumulative COVID-19 cases and its incidence, depend on time-independent variables. Estimating the parameters and model validation can be evaluated through the use of mathematical and statistical methods. The following forecasting and mathematical models with time-independent variables have been proposed by several researchers for describing and controlling COVID-19 outbreaks.

Forecasting models such as the logistic and Gompertz models have been adopted to forecast the total numbers of COVID-19 and severe acute respiratory syndrome (SARS) cases toward establishing a herd immunity policy [1–3]. The Bertalanffy model [4] and the Boltzmann growth curve [5, 6] have been used to describe the spread of COVID-19 and the waves of COVID-19 outbreaks. Regression analysis based on polynomial or ridge regression has been used to

---

establish models for forecasting the total numbers of cases and deaths during the COVID-19 pandemic [7, 8]. The ordinary least-squares estimator method with multiple regression analysis has been adopted to measure the impact of travel history on confirmed COVID-19 cases [9]. Logistic regression has been used to approximate the recovery probability of COVID-19 patients with respect to their demographic characteristics [10]. Forecasting models based on time series such as the (seasonal) autoregressive moving average (with an exogenous repressor) model, prophet models developed by using the time series concept, and machine learning models have been employed to study the trends of COVID-19 outbreaks such as the number of confirmed, recovered, and deaths have been established in many regions and countries toward establishing epidemiological control of the disease.

Decomposition in time series analysis of the confirmed numbers of cases and deaths in a COVID-19 outbreak has been used to identify seasonal trends and establish the mortality ratio and reproduction number of the disease [11–13]. Moreover, big data techniques, artificial intelligence procedures, data science, and machine learning algorithms have been applied to analyze COVID-19 outbreak data. For example, big data from the World Health Organization database, national databases, and social media communication databases has been utilized to forecast COVID-19 outbreaks toward establishing decision support systems for monitoring and controlling them [14, 15]. The predicted numbers of daily cases and deaths in various countries have been fitted with COVID-19 incidence data by using artificial neural networks (ANNs) [16]. The Gompertz model, the logistic model, the von Bertalanffy growth function, and the inverse ANN (IANN) have been used to forecast and compare the predictive accuracy with the actual number of cases, deaths, and recoveries during COVID-19 outbreaks [17]. The results show that the logistic model, Gompertz model, and IANN are the most suitable for forecasting the numbers of cases, deaths, and recoveries, respectively.

Mathematical models for describing the spreading characteristics of a COVID-19 outbreak have been applied to several aspects to help decision-makers implement relevant policies to control it. Model stability and estimating its parameters, such as the population size, birth rate, infection rate, and recovery rate, are important for forecasting the number of cases. These values provide fundamental information for decision-makers to instigate policies for monitoring, controlling, and protecting the majority of the population during a COVID-19 outbreak. The parameters of a mathematical model, such as quarantine rate, quarantine effectiveness, vaccination rate, vaccine efficacy, and rate of immunity loss, have been investigated to monitor the impact of quarantine and vaccination on the control of COVID-19 [18]. A susceptible (S), exposed (E), infectious (I), quarantined (Q), confirmed (C), recovered (R), and the concentration of the virus in the environmental reservoir (W) (SEIQCRW) transmission model, along with transmission routes from the environment to humans and human to human, have been applied to the COVID-19 outbreak to identify suitable prevention measures [19]. A SIR transmission model has been utilized for decision-making on lockdowns during the COVID-19 pandemic [20]. Asymptotic stability analysis of disease-free and endemic equilibria using the SEIR transmission model has been applied to establish the influences of asymptomatic cases and reinfection [21]. SIR transmission explained through a logistic model has been adopted to describe and interpret a COVID-19 outbreak [22]. The reproduction number from the SEIR transmission model has been utilized to analyze short-term COVID-19 incidence data [23]. A SEIQRD transmission model (D is death) has been employed to investigate the transmission dynamics of the COVID-19 outbreak. The reproduction number was calculated using a next-generation matrix approach [24]. SEIRD transmission models have been developed and used as a predictive model for the spread and control of COVID-19 [25] and to analyze healthcare demand and capacity for predicting and forecasting the impact of local COVID-19 outbreaks [26]. A SIR transmission model has been used to compute the effects of social distancing interventions during the COVID-19 pandemic [27]. Moreover, a SIRD compartmental model and variational autoencoder neural network have been used to forecast a COVID-19 pandemic in the future from the available historical big data on the subject [28].

As the above literature reviews, the models for describing the COVID-19 outbreak are based on time independent variable. For example, the forecasting models are the models with time independent variables such as the logistic and Gompertz models [1–3], the Bertalanffy model [4], the Boltzmann growth curve [5, 6], and Regression analysis [7, 8]. The mathematical models, a system of differential equations, are the models with time independent variable such SEIQCRW model [19], SIR model [20], SEIR model [21], and SEIQRD [24]. Most of these previous models are based on using time as the independent variable and do not involve distance and probabilistic analysis. Therefore, this research was undertaken to develop a model for describing the spread of a COVID-19 outbreak in which the computation for the models is based on probabilistic analysis by considering the distance from the center point of the outbreak. Applying a statistical control chart for monitoring and detecting the COVID-19 outbreak was also investigated. In addition, this research is applied to a case study of a COVID-19 outbreak at the Central Shrimp Market in Samut Sakhon Province, Thailand, as the center of a COVID-19 outbreak in the second wave.

2- Materials and Methods

Here, background on the spread of COVID-19 in Thailand and data collection for the probabilistic analysis of the effect of distance on the spread of COVID-19 are presented. In addition, the section contains detailed explanations of how the expected value, variance, and probabilistic analysis are derived.

2-1- Background on the COVID-19 Outbreak Depending on Distance in Thailand

The overall spread of COVID-19 has often been increased by outbreaks starting at a populous initial point of contact, such as a nightclub or market. The intensity of the viral spread has varied depending on the distance from the initial point of contact. For example, the COVID-19 outbreak at the Central Shrimp Market in Samut Sakhon province on December 23, 2020, spread as illustrated in Figure 1.
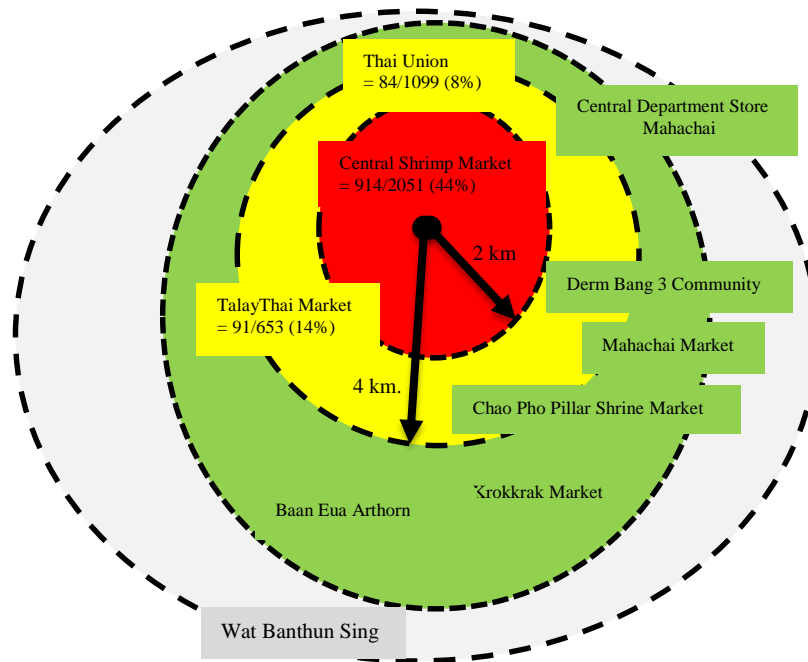
Figure 1. COVID-19 spread based on distance in Thailand. Source: the COVID-19 information center of Thailand and the Center for COVID-19 Situation Administration (CCSA) of Thailand

At the Central Shrimp Market in Samut Sakhon province, a 67-year-old woman was identified as the first case, and hence, the point of origin of the COVID-19 second wave, which began on December 17, 2020. The virus then spread quickly to the surrounding area albeit non-uniformly. The number of positive cases as a percentage of the population tested depended on the distance from the initial point of contact. Soon after the outbreak had been identified, testing was carried out in the surrounding area comprising 2,051 samples within 2 km of the initial point of contact, of which 914 (44%) were positive, and 2,272 samples within 4 km, of which 271 (11.93%) were positive. Nine hundred and ninety samples were taken in areas that were more than 4 km from the initial point of contact, with 20 positive results (2.02%). It is clear that the distance from the initial point of contact is related to the total number of COVID-19 cases identified in this outbreak. The present research is based on that clear relationship; it is designed to study the probability of a COVID-19 outbreak depending on the distance from the outbreak center. Differential equations and probabilistic analysis are adopted to analyze the data for the COVID-19 second-wave outbreak in Thailand. The spread of the COVID-19 outbreak in Samut Sakhon province can be estimated by using the concept of the derivatives of the total number of COVID-19 cases [29]. It can be seen in Figure 2 that the peak of the outbreak was around January 24, 2021.
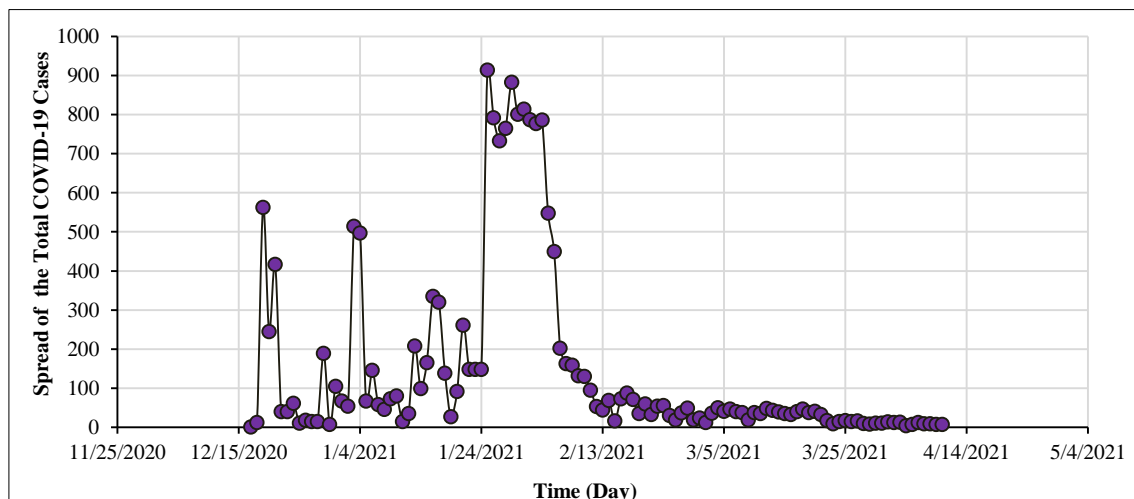

Figure 2. The estimated spreading from the COVID-19 outbreak in Samut Sakhon

2-2- Data Collection

Data and samples for this research consist of both real and simulated data. The real data and samples for this research comprise the total number of COVID-19 cases linked to the Central Shrimp Market in Samut Sakhon, Thailand, while the simulated data are the results of mathematical and statistical calculations. Data regarding the total number of COVID-19 cases for this research were collected in two aspects:

- Before the peak outbreak stage at around one month after the first outbreak (around January 3, 2021)

- The peak outbreak stage (around January 24, 2021)

- After the peak outbreak stage (around February 12, 2021), which is the day with the maximum number of daily new cases and the decrease in this number thereafter.

The total number of COVID-19 cases and distances (unit: km) between two positions in Thailand are used as the observations for this research. The second wave of the COVID-19 virus began with an outbreak that spread around the Central Shrimp Market in Samut Sakhon province, Thailand, where a super spreading event began on 17 December, 2020. The total number of COVID-19 cases was gathered from Worldometer [30], where COVID-19 data such as the total number of COVID-19 cases, new cases, and total deaths are updated daily. The distances from Samut Sakhon province to the other provinces were gathered from the DistanceFromTo website [31], as shown in Figure 3. For example, the distance from Samut Sakhon to Bangkok is approximately 33.69 km, the distance from Samut Sakhon to Nakhon Pathom is 38.62 km, and so on. The total number of COVID-19 cases was gathered from December 18, 2020, to April 01, 2021. Moreover, the indexes for the distances were ranked from the shortest to the farthest with the starting point at Samut Sakhon (the center of the outbreak). For example, index 1 refers to Samut Sakhon, index 2 refers to Samut Songkhram, index 3 refers to Bangkok, and so on.
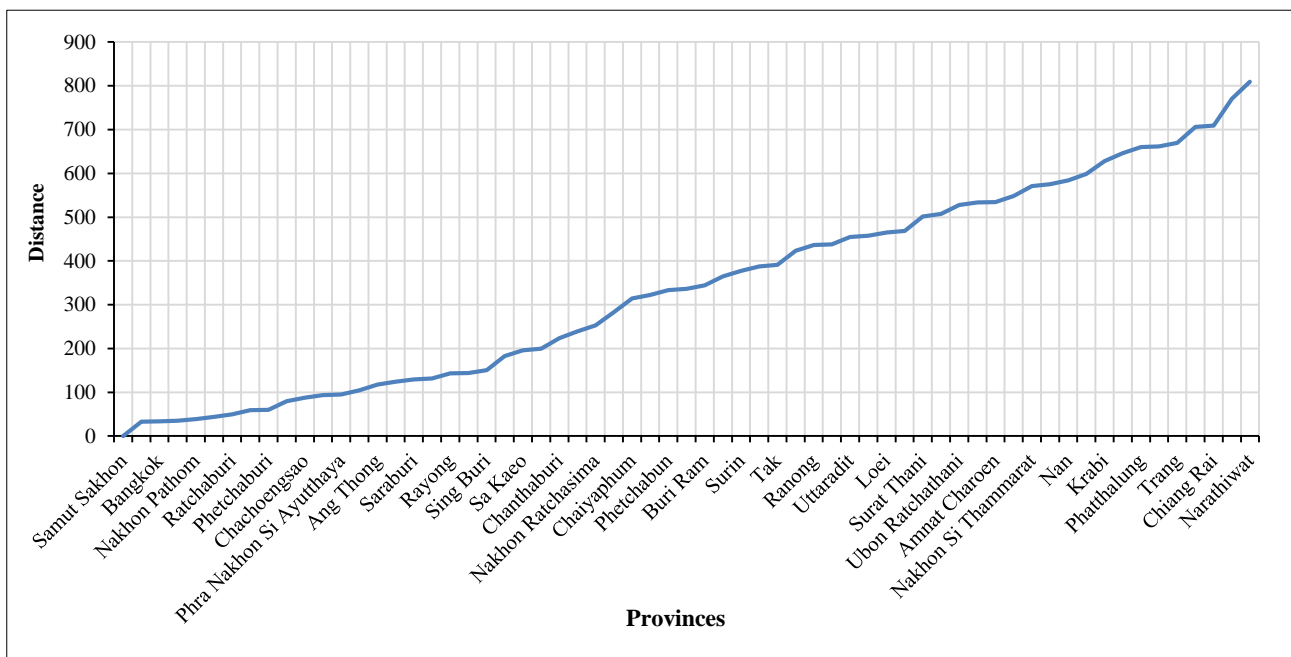
Figure 3. Distances from Samut Sakhon province to the other provinces in Thailand

2-3- The Forecasting Model for the COVID-19 Outbreak Depending on Distance

The assumption in this research is that the distance from the centre of the COVID-19 outbreak to other points is related to the total number of COVID-19 cases. Namely, the total number of COVID-19 cases varies with the distance from the outbreak centre. The key question to answer is: How will the COVID-19 infection rate change according to the distance from the outbreak centre? Let p(s) be the total number of COVID-19 cases at each distance. The differential equation is based on the assumption that the rate of change in the total number of COVID-19 cases is directly related to the distance from the initial contact point; i.e.,

$$\frac{dp}{ds} = kp \tag{1}$$

where k is the growth rate of p with respect to distance s.

When the method of separation is applied to Equation 1, the result becomes

$$\int \frac{dp}{p} = \int k \, ds$$

$ln|p| = ks + C_1$ where C_1 is an arbitrary constant

$$\exp(\ln|p|) = \exp(ks + C_1)$$

$$|p| = \exp(ks)\exp(C_1)$$

P=Cexp(ks), where C is $\pm\exp(C_1)$.

Therefore, the forecasting model for describing the total number of COVID-19 cases depending on distance m is:

$$p(s) = \begin{cases} 0 & ; s \in (m, \infty) \\ C \exp(ks); & s \in [0, m], \ m = \max(s) \\ 0 & ; s \in (-\infty, 0) \end{cases} \tag{2}$$

2-4- Probability Analysis of the COVID-19 Outbreak Depending on the Distance from the Initial Contact Point

The probability density function (PDF) for the distance from the initial contact point s, which is a random variable, corresponds to the property of probability $\int_{-\infty}^{+\infty} p(s)d(s) = 1$ such that

$$C = \frac{k}{\exp(km) - 1}$$

Therefore, the PDF for the assumption that the total number of COVID-19 cases will be the same at distance s is:

$$pr(s) = \frac{k}{\exp(km) - 1}\exp(ks); k \in \text{Real}, s \in [0, m], \ m = \max(s) \tag{3}$$

The area to the left of 0 can be derived as $0 \leq s \leq m$ can be drived as

$$P(S < s) = \int_0^s pr(t)dt = \int_0^s \frac{k}{\exp(km) - 1}\exp(kt)dt = \frac{k}{\exp(km) - 1}\int_0^s \exp(kt)dt = \frac{\exp(ks) - 1}{\exp(km) - 1}$$

Likewise, the area to the right of $0 \leq s \leq m$ can be drived as;

$$P(S > s) = \int_s^m pr(t)dt = \int_s^m \frac{k}{\exp(km) - 1}\exp(kt)dt = \frac{k}{\exp(km) - 1}\int_s^m \exp(kt)dt = \frac{\exp(km) - \exp(ks)}{\exp(km) - 1}$$

The area between distances a and b, where $(a, b) \subseteq (0, m)$; m=max(s), can be drivesd as

$$P(a \leq S \leq b) = \int_a^b pr(t)dt = \int_a^b \frac{k}{\exp(km) - 1}\exp(kt)dt = \frac{k}{\exp(km) - 1}\int_a^b \exp(kt)dt = \frac{\exp(kb) - \exp(ka)}{\exp(km) - 1}; b > a \tag{4}$$

The conditional probability that the total number of COVID-19 cases will be the same at a distance further away than b km when already knowing that the distance is more than a km (b>a) can be derived as

$$P(S > b \mid S > a) = \frac{P(S > b \cap S > a)}{P(S > a)} = \frac{P(S > b)}{P(S > a)} = \frac{\exp(km) - \exp(kb)}{\exp(km) - \exp(ka)}; b > a \tag{5}$$

Next, deriving the cumulative distribution function (CDF), $F(s) = \int_{-\infty}^s pr(x)dx$, for the total number of COVID-19 cases at a certain distance is conducted as follows.

Let S be the random variable for the distance from the COVID-19 outbreak. Its relationship is as follows:

$$F(s) = \int_{-\infty}^s pr(x)dx = \int_0^s \frac{k}{\exp(km) - 1}\exp(kx)dx = \frac{k}{\exp(km) - 1}\left[\frac{\exp(kx)}{k}\right]_{x=0}^{x=s} = \frac{\exp(ks) - 1}{\exp(km) - 1}$$

Therefore, the CDF for the total number of COVID-19 cases according to distance is

$$F(s) = \frac{\exp(ks) - 1}{\exp(km) - 1}; k \in \text{Real}, s \in [0, m], \ m = \max(s) \tag{6}$$

Next, the quantile function (QF or Q_p) and the median distance for the total number of COVID-19 cases according to distance are derived as follows.

The QF: $Q_p = F^{-1}(p) = \min\{s \in \text{Real}; F(s) \geq p\}; p \in (0,1)$

$$p = F(s) = \frac{\exp(ks) - 1}{\exp(km) - 1}$$

$$p\big(\exp(km) - 1\big) = \exp(ks) - 1$$

$$\exp(ks) = p\big(\exp(km) - 1\big) + 1$$

$$s = \frac{1}{k}\ln\big[\,p\big(\exp(km) - 1\big) + 1\big] = F^{-1}(p) = Q_p \tag{7}$$

For example:

$$Q_{0.01} = F^{-1}(0.01) = \frac{1}{k}\ln\big[\,0.01\big(\exp(km) - 1\big) + 1\big], \; Q_{0.25} = F^{-1}(0.25) = \frac{1}{k}\ln\big[\,0.25\big(\exp(km) - 1\big) + 1\big],$$

$$Q_{0.5} = F^{-1}(0.5) = \frac{1}{k}\ln\big[\,0.5\big(\exp(km) - 1\big) + 1\big], \; Q_{0.75} = F^{-1}(0.75) = \frac{1}{k}\ln\big[\,0.75\big(\exp(km) - 1\big) + 1\big],$$

$$Q_{0.99} = F^{-1}(0.99) = \frac{1}{k}\ln\big[\,0.99\big(\exp(km) - 1\big) + 1\big]$$

The median distance is the value of s for which F(s)=0.5:

$$\frac{\exp(ks) - 1}{\exp(km) - 1} = \frac{1}{2}$$

$$\exp(ks) - 1 = \frac{\exp(km) - 1}{2}$$

$$\exp(ks) = \frac{\exp(km) - 1}{2} + 1$$

$$s_{median} = \frac{1}{k}\ln\left(\frac{\exp(km) - 1}{2} + 1\right)$$

Therefore, the quantile function is $Q_p = F^{-1}(p) = \frac{1}{k}ln[p(exp(km) - 1) + 1]$ and the median is

$$s_{median} = \frac{1}{k}\ln\left(\frac{\exp(km) - 1}{2} + 1\right)$$

The properties of the PDF such as the expected value (sometimes called the first moment) and the variance, among others, are defined as follows.

The first moment (the expected value or the mean of the population):

$$E(S) = \int_0^m spr(s)ds = \int_0^m s\frac{k}{\exp(km) - 1}\exp(ks)ds = \frac{k}{\exp(km) - 1}\left(\frac{s\exp(ks)}{k} - \frac{\exp(ks)}{k^2}\right)_{s=0}^{s=m}$$

$$= \frac{k}{\exp(km) - 1}\left(\frac{m\exp(km)}{k} - \frac{\exp(km)}{k^2} + \frac{1}{k^2}\right) = \frac{1}{\exp(km) - 1}\left(m\exp(km) + \frac{1 - \exp(km)}{k}\right) \tag{8}$$

The second moment:

$$E(S^2) = \int_0^m s^2 pr(s)ds = \int_0^m s^2 \frac{k}{\exp(km) - 1}\exp(ks)ds$$

$$= \frac{k}{\exp(km) - 1}\int_0^m s^2 \exp(ks)ds$$

$$= \frac{k}{\exp(km) - 1}\left[\exp(ks)\left(\frac{s^2}{k} - \frac{2s}{k^2} + \frac{2}{k^3}\right)\right]_{s=0}^{s=m}$$

$$= \frac{k}{\exp(km) - 1}\left[\exp(km)\left(\frac{m^2}{k} - \frac{2m}{k^2} + \frac{2}{k^3}\right) - \frac{2}{k^3}\right]$$

$$= \frac{k}{\exp(km) - 1}\left[\exp(km)\left(m^2 - \frac{2m}{k} + \frac{2}{k^2}\right) - \frac{2}{k^2}\right]$$

The third moment:

$$E(S^3) = \int_0^m s^3 \, pr(s)ds = \int_0^m s^3 \frac{k}{\exp(km)-1}\exp(ks)ds = \frac{k\exp(km)}{\exp(km)-1}\left[\frac{6}{k^3}m - \frac{3}{k^2}m^2 + \frac{1}{k}m^3\right]$$

The fourth moment:

$$E(S^4) = \int_0^m s^4 \, pr(s)ds = \int_0^m s^4 \frac{k}{\exp(km)-1}\exp(ks)ds = \frac{k\exp(km)}{\exp(km)-1}\left[-\frac{24}{k^4}m + \frac{12}{k^3}m^2 - \frac{4}{k^2}m^3 + \frac{1}{k}m^4\right]$$

Variance:

$$Var(S) = E(S^2) - [E(S)]^2 = \frac{k}{\exp(km)-1}\left[\exp(km)\left(m^2 - \frac{2m}{k} + \frac{2}{k^2}\right) - \frac{2}{k^2}\right] - \left[\frac{1}{\exp(km)-1}\left(m\exp(km) + \frac{1-\exp(km)}{k}\right)\right]^2$$

Skewness:

$$Skew(S) = E\left[\left(\frac{S-E(S)}{Var(S)}\right)^3\right] = \frac{E(S^3) - 3E(S)E(S^2) + 2[E(S)]^3}{[Var(S)]^{\frac{3}{2}}} = \frac{E(S^3) - 3E(S)Var(S) - [E(S)]^3}{[Var(S)]^{\frac{3}{2}}}$$

$$= \frac{\begin{aligned}&\left(\frac{k\exp(km)}{\exp(km)-1}\left[\frac{6}{k^3}m - \frac{3}{k^2}m^2 + \frac{1}{k}m^3\right]\right) - 3\left[\frac{1}{\exp(km)-1}\left(m\exp(km) + \frac{1-\exp(km)}{k}\right)\right]\\ &\times\left(\frac{k}{\exp(km)-1}\left[\exp(km)\left(m^2 - \frac{2m}{k} + \frac{2}{k^2}\right) - \frac{2}{k^2}\right] - \left[\frac{1}{\exp(km)-1}\left(m\exp(km) + \frac{1-\exp(km)}{k}\right)\right]^2\right)\\ &-\left[\frac{1}{\exp(km)-1}\left(m\exp(km) + \frac{1-\exp(km)}{k}\right)\right]^3\end{aligned}}{\left(\frac{k}{\exp(km)-1}\left[\exp(km)\left(m^2 - \frac{2m}{k} + \frac{2}{k^2}\right) - \frac{2}{k^2}\right] - \left[\frac{1}{\exp(km)-1}\left(m\exp(km) + \frac{1-\exp(km)}{k}\right)\right]^2\right)^{\frac{3}{2}}}$$

Kurtosis:

$$Kurt(S) = E\left[\left(\frac{S-E(S)}{Var(S)}\right)^4\right]$$

$$= \frac{E(S^4) - 4E(S)E(S^3) + 6[E(S)]^2 E(S^2) - 3[E(S)]^4}{[Var(S)]^2}$$

$$= \frac{E(S^4) - 4E(S)E(S^3) + 6[E(S)]^2 E(S^2) + 3[E(S)]^4}{[Var(S)]^2}$$

$$= \frac{\begin{aligned}&\left(\frac{k\exp(km)}{\exp(km)-1}\left[-\frac{24}{k^4}m + \frac{12}{k^3}m^2 - \frac{4}{k^2}m^3 + \frac{1}{k}m^4\right]\right)\\ &-4\left[\frac{1}{\exp(km)-1}\left(m\exp(km) + \frac{1-\exp(km)}{k}\right)\right]\left(\frac{k\exp(km)}{\exp(km)-1}\left[\frac{6}{k^3}m - \frac{3}{k^2}m^2 + \frac{1}{k}m^3\right]\right)\\ &+6\left[\frac{1}{\exp(km)-1}\left(m\exp(km) + \frac{1-\exp(km)}{k}\right)\right]^2\left(\frac{k}{\exp(km)-1}\left[\exp(km)\left(m^2 - \frac{2m}{k} + \frac{2}{k^2}\right) - \frac{2}{k^2}\right]\right)\\ &+3\left(\left[\frac{1}{\exp(km)-1}\left(m\exp(km) + \frac{1-\exp(km)}{k}\right)\right]\right)^4\end{aligned}}{\left(\frac{k}{\exp(km)-1}\left[\exp(km)\left(m^2 - \frac{2m}{k} + \frac{2}{k^2}\right) - \frac{2}{k^2}\right] - \left[\frac{1}{\exp(km)-1}\left(m\exp(km) + \frac{1-\exp(km)}{k}\right)\right]^2\right)^2}$$

Skewness and kurtosis are evaluated and interpreted to measure dispersion. Kurtosis is used to measure the peakedness of the PDF. Meanwhile, if the skewness is approximately zero, then the PDF is symmetrical, whereas if it is negative or positive, then it is not. Therefore, if kurt(s) is approximately zero, negative, or positive then the PDF is a symmetric, platykurtic, or leptokurtic (heavy-tailed) function, respectively

Next, the sojourn function (SF) denoted as j(s) is the probability that the total number of COVID-19 cases still exist after distance s, while the hazard function (HF) denoted as h(s) is the probability that if the total number of COVID-19 cases still exist after distance s, then the total number of COVID-19 cases that do not exist after distance s is s+Δs, where Δs is small. Thus

$$j(s) = pr(S > s) = \int_s^m p(x)dx = 1 - pr(S \le s) = 1 - \frac{\exp(ks) - 1}{\exp(km) - 1} \qquad (9)$$

$$h(s) = pr(s \le S < s + \Delta s \mid S > s)$$

$$= \lim_{\Delta s \to 0} \frac{j(s) - j(s + \Delta s)}{\Delta s}$$

$$= -\lim_{\Delta s \to 0} \frac{j(s + \Delta s) - j(s)}{\Delta s}$$

$$= -\frac{d}{ds} j(s)$$

$$= -\frac{d}{ds}\left(1 - \frac{\exp(ks) - 1}{\exp(km) - 1}\right)$$

$$= \frac{\exp(ks)}{k\left[\exp(km) - 1\right]} \qquad (10)$$

2-5- Analysis of the PDF

The PDF of the total number of COVID-19 cases, which relies on the growth rate (k) and the distance (m), can be categorized into three cases.

Theorem 1: The PDF denoted by $pr_{k,m}(s)$ for distance m of the total number of COVID-19 cases relies on k and m.

Case 1: if k>0 and for all $\{m_1 \le m_2\} \in Max(s)$, then $pr_{k,m_1}(s) \ge pr_{k,m_2}(s)$.

Case 2: if k<0 and for all $\{m_1 \le m_2\} \in Max(s)$, then $pr_{k,m_1}(s) \ge pr_{k,m_2}(s)$.

Case 3: if k=0 and for all $\{m_1 \le m_2\} \in Max(s)$, then the PDF for the distance of the total number of COVID-19 cases does not exist.

Therefore, if $\{m_1 \le m_2\} \in Max(s)$, then the PDF that the total number of COVID-19 cases which will remain at distance m_1 is more than m_2.

Proof.

Let r be a positive real number.

For all $m_1, m_2 \in Max(s)$ such that $m_1 \le m_2$, then

$$\frac{1}{\exp(rm_1) - 1} \ge \frac{1}{\exp(rm_2) - 1},$$

$$\frac{r\exp(rs)}{\exp(rm_1) - 1} \ge \frac{r\exp(rs)}{\exp(rm_2) - 1},$$

$$pr_{r,m_1}(s) \ge pr_{r,m_2}(s).$$

Thus case 1 holds.

Let n be a negative real number.

For all $m_1, m_2 \in Max(s)$ such that $m_1 \le m_2$, then

$$\frac{1}{\exp(nm_1) - 1} \le \frac{1}{\exp(nm_2) - 1},$$

$$\frac{n\exp(ns)}{\exp(nm_1) - 1} \ge \frac{n\exp(ns)}{\exp(nm_2) - 1},$$

$$pr_{n,m_1}(s) \ge pr_{n,m_2}(s).$$

Thus, case 2 holds.

Obviously, by using the monotonicity property of the Riemann integral $\int_0^m pr_{k,m_1}(s)ds \geq \int_0^m pr_{k,m_2}(s)ds$, then case 3 holds.

2-6- Estimation of the Forecasting Model Parameters and Model Comparison

Parameter estimation is performed by using least-squares-error estimation. The error between the actual value and forecasted value is used to define the sum-of-squares error (SSE) function as follows:

$$SSE(C,k) = (pa - pf)^2 \tag{11}$$

where pa and pf are the actual and forecasted values of the total number of COVID-19 cases, respectively.

To minimize the SSE, the critical value of the SSE function is determined by using the first partial derivative with respect to parameters C and k, which is zero. Therefore, we can respectively define

$\frac{\partial}{\partial C}SSE(C,k) = 0$ and $\frac{\partial}{\partial k}SSE(C,k) = 0$.

Let $D(C^*, k^*)$ be the determinant of the partial derivatives of the SSE function such that

$$D(C^*,k^*) = \begin{vmatrix} SSE_{cc}(C^*,k^*) & SSE_{ck}(C^*,k^*) \\ SSE_{kc}(C^*,k^*) & SSE_{kk}(C^*,k^*) \end{vmatrix} \tag{12}$$

Optimal values C^*, k^* will occur at $D(C^*, k^*) > 0$ and $SSE_{cc}(C^*, k^*) > 0$ with respect to the minimum of the SSE function at the critical point. However, the algorithm for solving the parameter estimations is based on the Levenberge-Marquardt Algorithm [32, 33].

For measuring the accuracy of the parameter estimations, we use the root-mean-squared error (RMSPE), which is a measurement of the error, and the coefficient of determination (R^2), which is a measurement of how much independent variable s can explain dependent variable p. These are respectively computed as

$$RMSPE = \sqrt{\frac{1}{m}\sum_{s=1}^{m}\left(\frac{pa(s) - pf(s)}{pa(s)}\right)^2} \tag{13}$$

where pa(s) and pf(s) are the actual and forecasted values of the total number of COVID-19 cases at any distance s, respectively, and

$$R^2 = \frac{\text{Explained variance}}{\text{Total variance}} \tag{14}$$

where the total variance is the sum of the squares of the differences between the actual values and the mean of the actual values (i.e., total variance is $\sum_{s=1}^{m}(pa(s) - \overline{pa}(s))^2$ and the explained variance is the sum of the squares of the differences between the forecasted values and the mean of the actual values; i.e., explained variance is $\sum_{s=1}^{m}(pf(s) - \overline{pa}(s))^2$. For the model comparison, the best performing one has the lowest RMSPE value and the highest R^2 value.

2-7- Application of a Control Chart for Monitoring COVID-19 Outbreaks

In this research, the exponentially weighted moving average (EWMA) control chart is applied to monitor COVID-19 outbreaks [34]. To construct the EWMA control chart, the methods for estimation of the expected value and variance of the total number of COVID-19 cases can be calculated by using two methods as follows.

- The delta method [35]:

This is based on a Taylor series expansion about independent distance variable s=E(S) for the total number of COVID-19 cases represented by function p=Cexp(ks); i. e. the expected value of p can be expressed as:

$$E(p) = p(E(S)) + \frac{1}{2}\frac{d^2p}{ds^2}(E(S))Var(S) + \frac{1}{6}\frac{d^3p}{ds^3}(E(S))Var(S)E\left[(s - E(S))^3\right] + ...$$

and the variance of p can be expressed as:

$$Var(p) = \sigma_p^2 = Var\left[p(E(S)) + (p - E(S))\frac{dp}{ds}(E(S))\right] = \left[\frac{d^2p}{d^2s}(E(S))\right]^2 Var(S) + ...$$

The estimated expected value of p by using the second derivative $\frac{d^2p}{ds^2} = Ck^2\exp(ks)$ is given by:

$$E(p) \approx p(E(S)) + \frac{1}{2}\frac{d^2 p}{ds^2}(E(S))Var(S) \tag{15}$$

and the estimated variance of p by using the second derivative $\frac{d^2 p}{ds^2} = Ck^2 \exp(ks)$ is given by:

$$Var(p) = \sigma_p^2 \approx \left[\frac{d^2 p}{d^2 s}(E(S))\right]^2 Var(S) \tag{16}$$

- The sample method:

This is based on the sample mean and variance of the total number of COVID-19 cases for sample size n.

The estimated expected value of p with index (i) for the distance is given by:

$$E(p) \approx \bar{p} = \frac{\sum_{i=1}^{n} p(s(i))}{n} \tag{17}$$

and the estimated variance of p with index (i) for the distance is given by:

$$Var(p) \approx \sigma_p^2 \approx \frac{\sum_{i=1}^{n} \left(p(s(i) - \bar{p})\right)^2}{n-1} \tag{18}$$

Next, the EWMA control chart is applied for monitoring the total number of COVID-19 cases p(s(i)), which are independent and identically distributed random variables.

The EWMA control chart derived from EWMA statistic y with smoothing parameter $\lambda \in (0,1]$ can be defined as:

$$y(i) = \lambda p(i) + (1-\lambda)y(i-1); \; i = 1,2,3,...,n; \; y(0) = E(p) \tag{19}$$

where i is the index sequence for the distance.

For i=1, 2, 3, … , n, the expected value E(y(i)) of EWMA statistic y becomes:

$$E(y(i)) = \lambda E(p) + (1-\lambda)E(y(i-1)) = E(p)$$

For i=1, 2, 3, … , n, the variance $\sigma_{y(i)}^2$ of EWMA statistic y becomes:

$$\sigma_{y(i)}^2 = \frac{\lambda}{2-\lambda}\left[1 - (1-\lambda)^{2i}\right]\sigma_p^2$$

Therefore, the three control limits for the EWMA control chart for =1, 2, 3, … , n can be defined as:

Upper Control Limit (UCL)=$E(y(i)) + d\sqrt{\frac{\lambda}{\lambda-2}[1 - (1-\lambda)^{2i}]\sigma_p^2}$,

Center Line (CL)=E(y(i)), and:

Upper Control Limit (UCL)=$E(y(i)) - d\sqrt{\frac{\lambda}{\lambda-2}[1 - (1-\lambda)^{2i}]\sigma_p^2}$

where d is the width of the control limit.

For monitoring the COVID-19 outbreak by using the EWMA control chart, the first passage time (FPT) [36], which is the first signal outside of the upper or lower control limit of the EWMA control chart. The FPT should be as large as possible to indicate a lower severity of a COVID-91 outbreak.

The FPT based on distance for monitoring the total number of COVID-19 cases which is a decreasing function that can be defined in terms of the infimum (inf) [37] as follows:

$$FPT = \inf\{i > 0 \mid y(i) < LCL\} \text{ for i=1, 2, 3, … , n.}$$
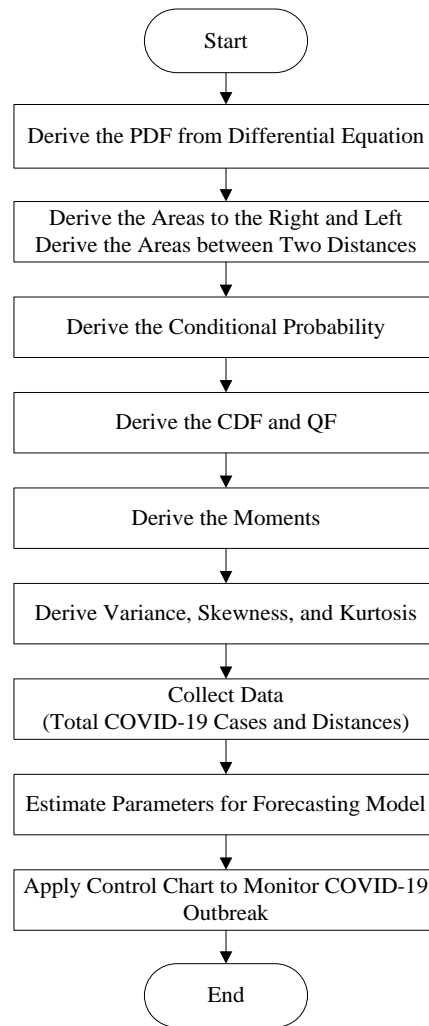
The FPT based on time (t) for monitoring the total number of COVID-19 cases which is an increasing function that can be defined in terms of the infimum as follows:

$$FPT = \inf\{t > 0 \mid y(t) > UCL\} \text{ for t=1, 2, 3, … , n}$$

The summarized process of the methodology is as follows (Figure 4).

```
                        ╭─────────────╮
                        │    Start    │
                        ╰─────────────╯
                               │
                               ▼
        ┌────────────────────────────────────────────┐
        │     Derive the PDF from Differential Equation │
        └────────────────────────────────────────────┘
                               │
                               ▼
        ┌────────────────────────────────────────────┐
        │  Derive the Areas to the Right and Left       │
        │  Derive the Areas between Two Distances       │
        └────────────────────────────────────────────┘
                               │
                               ▼
        ┌────────────────────────────────────────────┐
        │     Derive the Conditional Probability        │
        └────────────────────────────────────────────┘
                               │
                               ▼
        ┌────────────────────────────────────────────┐
        │          Derive the CDF and QF                │
        └────────────────────────────────────────────┘
                               │
                               ▼
        ┌────────────────────────────────────────────┐
        │           Derive the Moments                  │
        └────────────────────────────────────────────┘
                               │
                               ▼
        ┌────────────────────────────────────────────┐
        │   Derive Variance, Skewness, and Kurtosis     │
        └────────────────────────────────────────────┘
                               │
                               ▼
        ┌────────────────────────────────────────────┐
        │             Collect Data                      │
        │   (Total COVID-19 Cases and Distances)        │
        └────────────────────────────────────────────┘
                               │
                               ▼
        ┌────────────────────────────────────────────┐
        │  Estimate Parameters for Forecasting Model    │
        └────────────────────────────────────────────┘
                               │
                               ▼
        ┌────────────────────────────────────────────┐
        │  Apply Control Chart to Monitor COVID-19      │
        │              Outbreak                         │
        └────────────────────────────────────────────┘
                               │
                               ▼
                        ╭─────────────╮
                        │     End     │
                        ╰─────────────╯
```

Figure 4. The diagram of the research process

3- Results

Results from a simulation study and using the real data described in detail earlier are provided in this section. Table 1 reports the expected values, medians, variances, skewness, and kurtosis for the PDF, CDF, QF, SF, and HF for the total number of COVID-19 cases according to distance with parameters C, k, and m. Furthermore, the sample mean is 30 for positive k=0.5 and the sample mean is 100 for negative k=-0.75 (Figure 5 to 7). The 95% confidence interval (CI) for sample mean 30 in case of positive k=0.5 is (29.50, 30.50). The 95% confidence interval (CI) for sample mean 100 in case of positive k=-0.75 is (99.82, 100.18). With increasing distance from the COVID-19 outbreak, the PDF increased for positive k=0.5 (Figure 5-c)) and decreased for negative k=-0.75 (Figure 5-d)).

Table 1. The statistics for the simulated total number of COVID-19 cases outbreak data according to distance

| Statistic	Simulation Study Parameters	
	C=50, m=60, k=0.5>0	C=20, m=200, k=-0.75<0
Expected value	58.00	1.33
$Q_{0.01}, Q_{0.25}, Q_{0.5}, Q_{0.75}, Q_{0.99}$	50.79, 57.23, 58.61, 59.42, 59.98	0.01, 0.38, 0.92, 1.84, 6.14
Median	58.62	0.92
The 2nd moment	3.37e+03	3.56
The 3rd moment	1.96e+05	14.22
The 4th moment	1.14e+07	75.85
Variance	4.00	1.78
Skewness	-2.00	2.00
Kurtosis	4.2582e+06	84.00
Sample mean	30	100
95% CI	(29.50, 30.50)	(99.82, 100.18)

Note: CI, confidence interval for the sample mean; Q, quantile function; C, constant; m, the maximum distance; k, the growth rate.
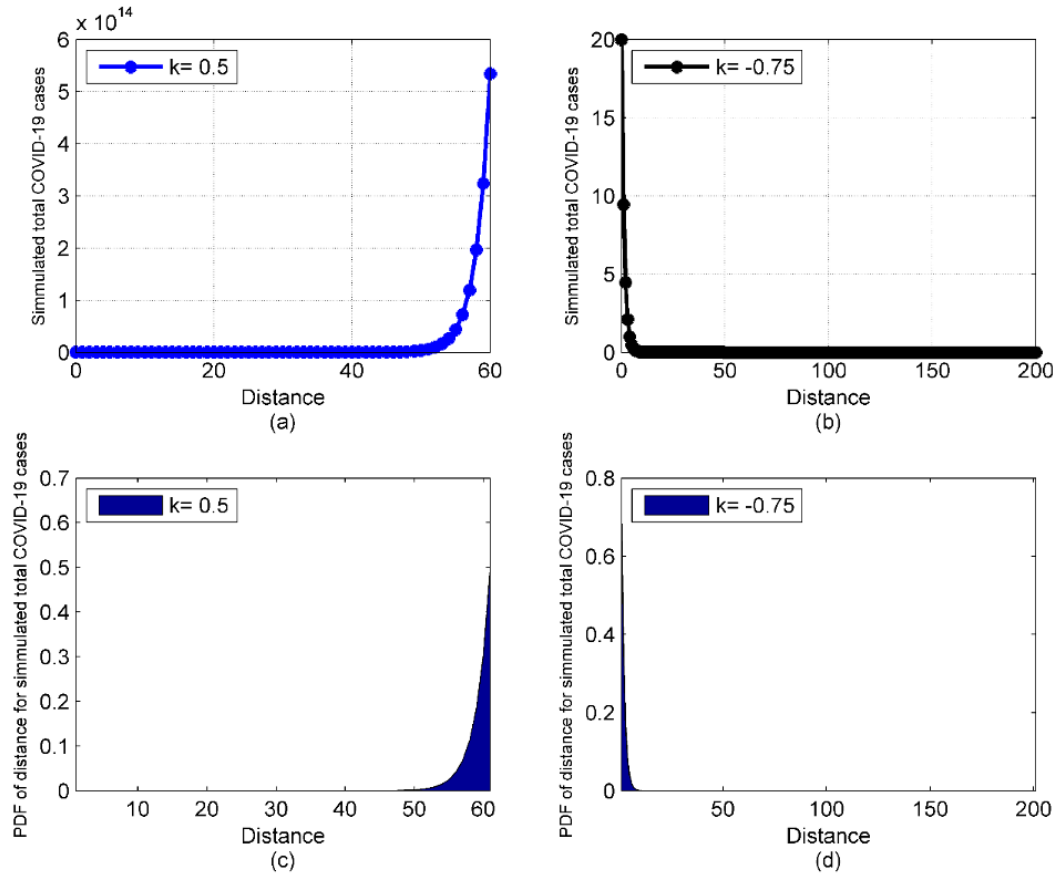
Figure 5. Simulated total number of COVID-10 cases based on distance (a) for positive k and (b) for negative k and the PDF for the simulated distance from the outbreak (c) for positive k and (d) for negative k
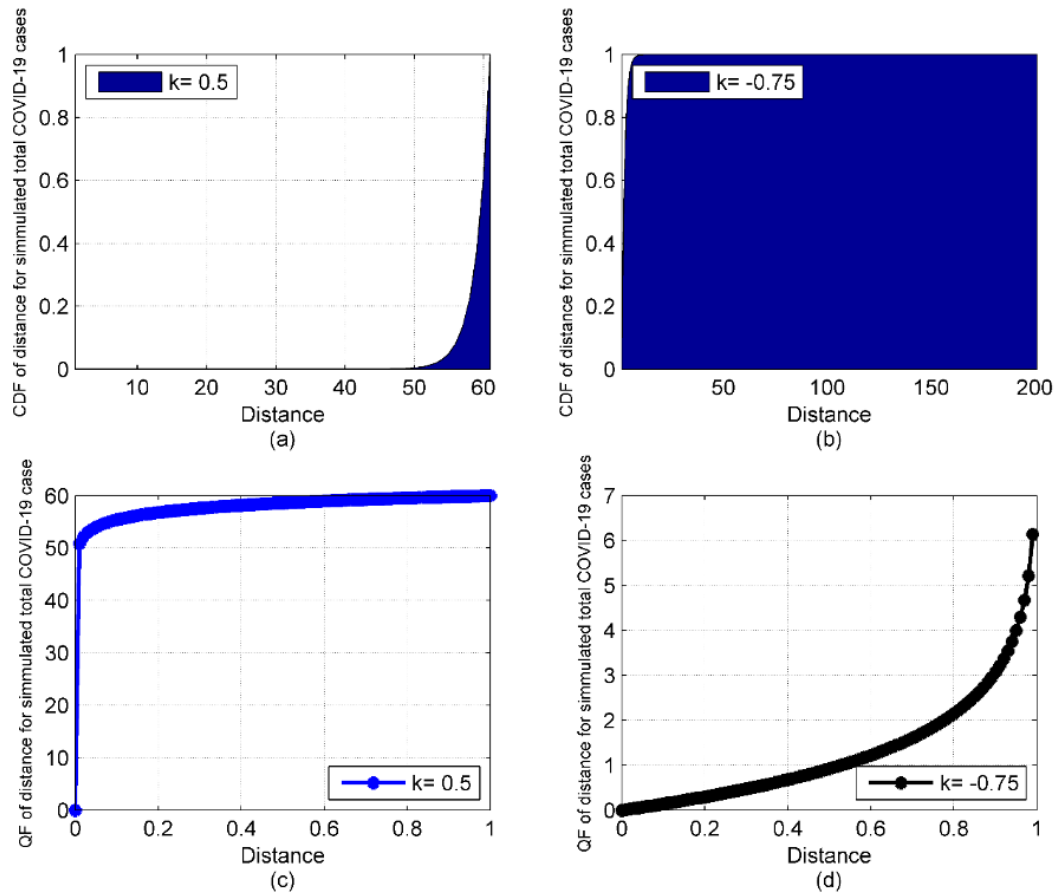
Figure 6. The (a) CDF and (c) QF for positive k and (b) CDF and (d) QF for negative k according to the distance from the COVID-19 outbreak with simulated data
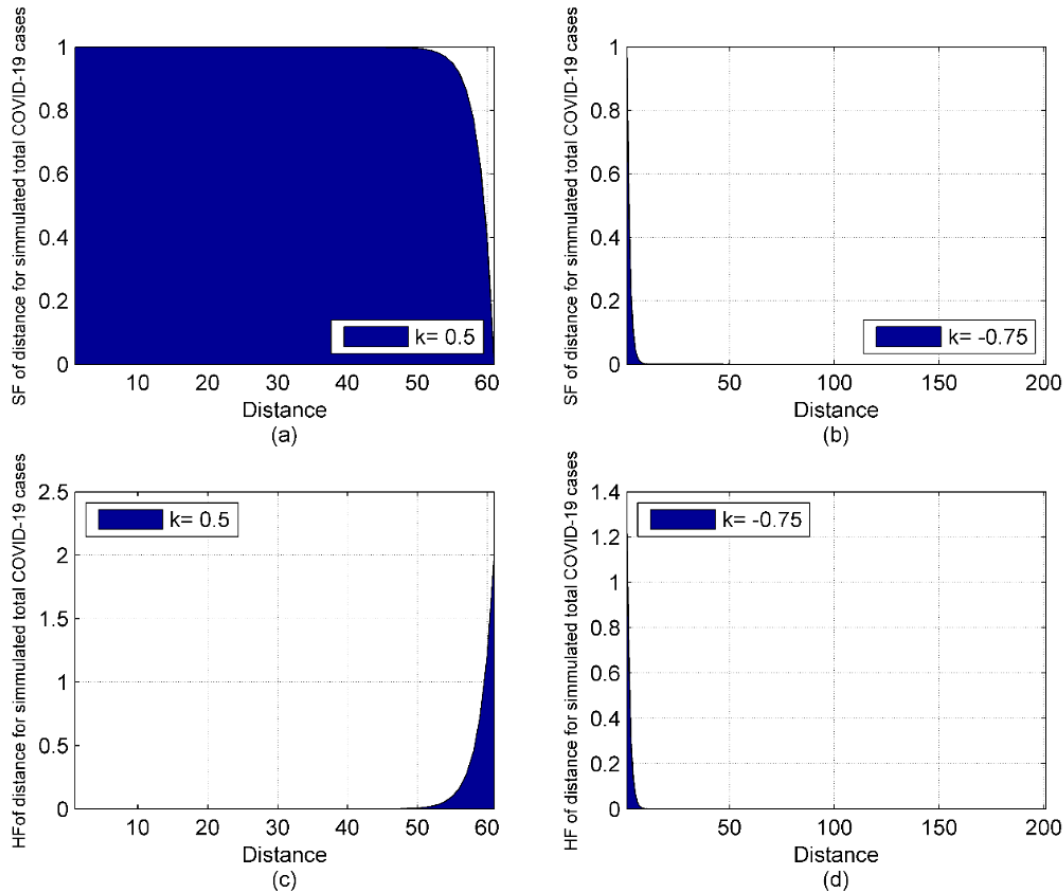
Figure 7. The (a) SF and (c) HF for positive k and (a) SF and (c) HF for positive k according to the distance from the COVID-19 outbreak with simulated data

Next, the results from the application of real data for the COVID-19 outbreak at the Central Shrimp Market in Samut Sakhon province are presented.

Table 2 reports the statistics estimated by using the model in Equation 2. For example, the expected values of the distance from the outbreak are 12.11, 12.09, and 13.22 for the before, peak, and after outbreak stages, respectively. The proportions of the variance for the total number of COVID-19 cases explained by the distance are measured by using R^2, which are 0.94, 0.97, and 0.99 for the before, peak, and after peak outbreak stages, respectively. For the error, it found that the RMSPE values are 1.92, 1.85, and 1.49 for the before peak, peak, and after peak outbreak stages, respectively. Meanwhile, the 95% CIs for sample mean 322.88 for the before peak outbreak, 339.50 for the peak outbreak, and 339.49 for the after peak outbreak are (319.54, 326.24), (336.51, 342.49), and (336.24, 342.76), respectively. In addition, estimated parameter k was negative in all three stages (-0.0825, -0.0826, and -0.0757 for the before peak, peak, for after peak outbreak stages, respectively). The estimated C are 1.86e+03, 5.33e+03, and 5.18e+03 for the before peak, peak, after peak outbreak stages, respectively. The variances are 175.31, 146.34, and 174.73 for the before peak, peak, after peak outbreak stages, respectively. The skewness is constant at 2.00 for all three stages. The kurtosis is constant at 84.00 for all three stages.

Table 2. The statistics for the PDF for the total number of COVID-19 cases according to the distance from the Central Shrimp Market in Samut Sakhon province

Statistic	COVID-19 Outbreak Stage		
	Before the Peak (Jan 3, 2021)	At the Peak (Jan 24, 2021)	After the Peak (Feb 12, 2021)
Estimated C	1.86e+03	5.33e+03	5.18e+03
Estimated k	-0.0825	-0.0826	-0.0757
m	809.61	809.61	809.61
RMSPE	1.92	1.85	1.49
R^2	0.94	0.97	0.99
Expected value	12.11	12.09	13.22
$Q_{0.01}, Q_{0.25}, Q_{0.5}, Q_{0.75}, Q_{0.99}$	0.12, 3.48, 8.40, 16.79, 55.78	0.12, 3.48, 8.39, 16.77, 55.71	0.13, 3.80, 9.16, 18.32, 60.87

Median	8.40	8.39	9.16
The 2nd moment	293.39	292.68	349.46
The 3rd moment	1.07e+04	1.06e+04	1.39e+04
The 4th moment	5.16e+05	5.14e+05	7.33e+05
Variance	175.31	146.34	174.73
Skewness	2.00	2.00	2.00
Kurtosis	84.00	84.00	84.00
Sample mean	322.88	339.50	339.49
95% CI	(319.54, 326.24)	(336.51, 342.49)	(336.24, 342.76)

Note: CI, confidence interval for the sample mean;Q, quantile function;C, constant; m, the maximum distance; k, the growth rate.

The estimated total number of COVID-19 cases for the Samut Sakhon COVID-19 outbreak with negative k for all three outbreak stages are shown in Figures 8-a to 8-c. The PDF values for the distance from the Samut Sakhon COVID-19 outbreak for all three outbreak stages are presented in Figures 8-d to 8-f. It can be seen that the PDF values are decreasing functions when the distance increases for negative k in all three stages.
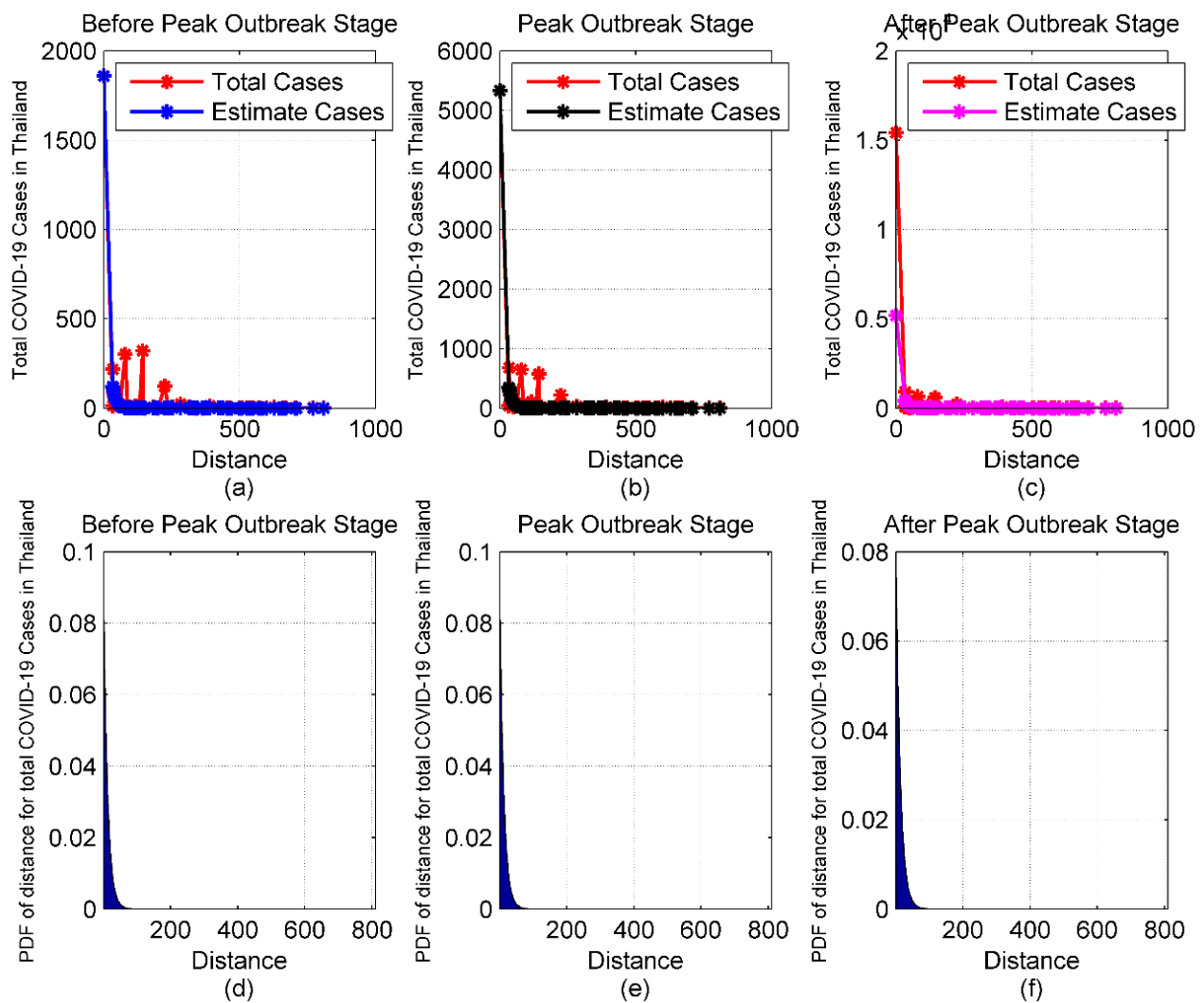
Figure 8. Estimated and real total numbers of COVID-19 cases and the PDF of the distance from the COVID-19 outbreak in Samut Sakhon for the (a,d) before peak outbreak stage, (b,e) peak outbreak stage, and (c,d) after peak outbreak stage, respectively.

Figures 9 shows the CDF and QF for the distance from the COVID-19 outbreak for all outbreak stages. It can be seen that the CDF converges to one when the distance increases, while the QF according to the probability increases within the interval 0 and 1. Figure 10 shows the SF and HF for the distance from the Samut Sakhon COVID-19 outbreak, in which it can be seen that both decrease with increasing distance.
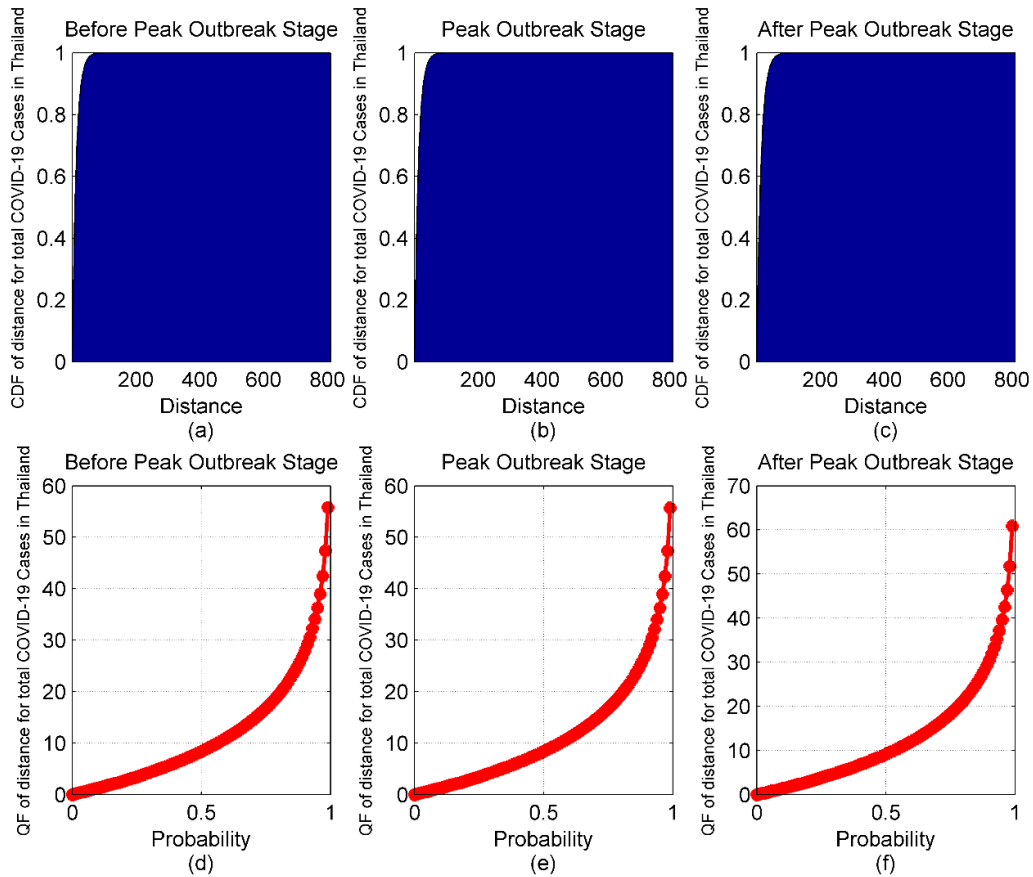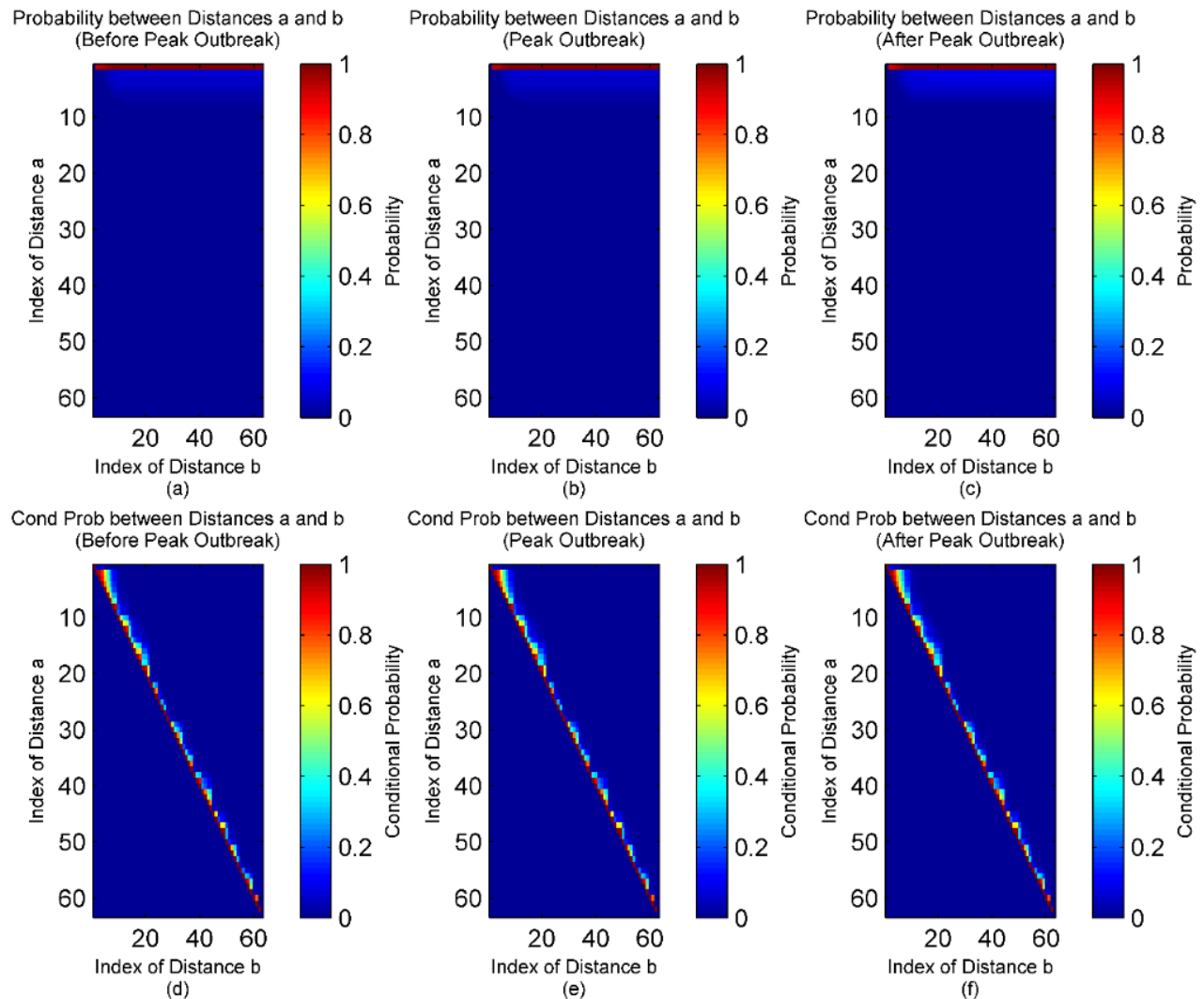
Figure 9. The CDF and QF for the distance from the Samut Sakhon COVID-19 outbreak for the (a,d) before peak outbreak stage, (b,e) peak outbreak stage, and (c,f) after peak outbreak stage, respectively
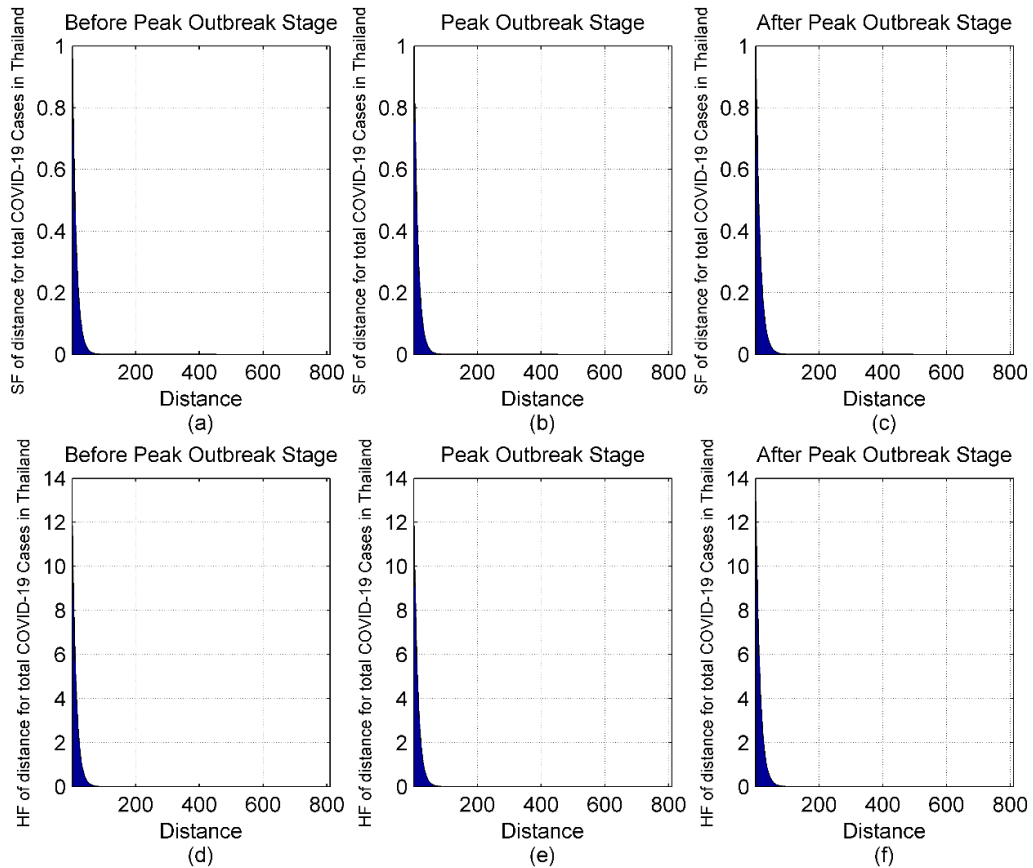

Figure 10. SF and HF according to the distance from the Samut Sakhon COVID-19 outbreak for the (a,d) before peak outbreak stage, (b,e) peak outbreak stage, and (c,f) after peak outbreak stage, respectively

Next, the probability and conditional probability between two distances a and b from the COVID-19 outbreak centre at Samut Sakhon in Thailand for 63 provinces are depicted in Figure 11. The probabilities that the COVID-19 outbreak will spread from Samut Sakhon (index 1: 0 km) to Bangkok (index 3: 33.69 km) are approximately 0.938, 0.938, and 0.922 for the before peak, peak, and after peak outbreak stages, respectively. Moreover, the conditional probabilities that the COVID-19 outbreak will continue after the distance to Samut Prakan (index 4: 35.35 km) are approximately 0.872, 0.872, and 0.882 for the before peak, peak, and after peak outbreak stages, respectively. Moreover, the probability and conditional probability decrease and increase after the peak outbreak stage, respectively. For example, the probability values between index of distance 3 and 1 are 0.938, 0.938, and 0.922 for before peak, peak, and after peak stage, respectively. The conditional probability values between index of distance 3 and 4 are 0.872, 0.872, 0.882 for before peak, peak, and after peak stage, respectively.


Figure 11. Probability and conditional probability between two distances from the Samut Sakhon COVID-19 outbreak for the (a,d) before peak outbreak stage, (b,e) peak outbreak stage, and (c,f) after peak outbreak stage, respectively

Next, applying the EWMA control chart for comparing the time and distance domains for the before peak, peak, and after peak outbreak stages of the Samut Sakhon COVID-19 outbreak are illustrated in Figures 12 to 17. The EWMA control charts using the delta method for estimating the expected values and the variances are illustrated in Figures 12-a to 17-a and Figures 12-b to 17-b, respectively, while those using the sample method are illustrated in Figures 12-c to 17-c and Figures 12-d to 17-d, respectively. A smoothing parameter (λ) value of 0.3 for the EWMA statistic is used in Figures 12-a to 17-a and Figures 12-c to 17-c while 0.7 is used in Figures 12-b to 17-b and Figures 12-d to 17-d.
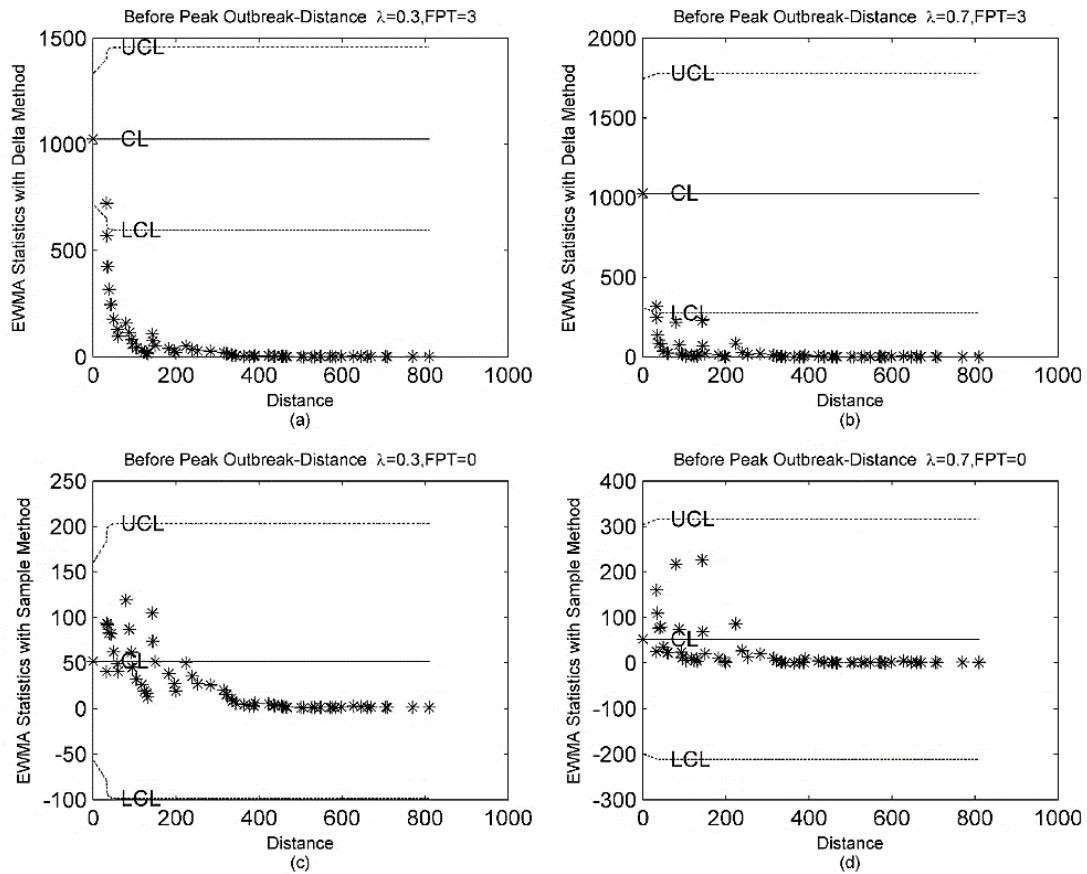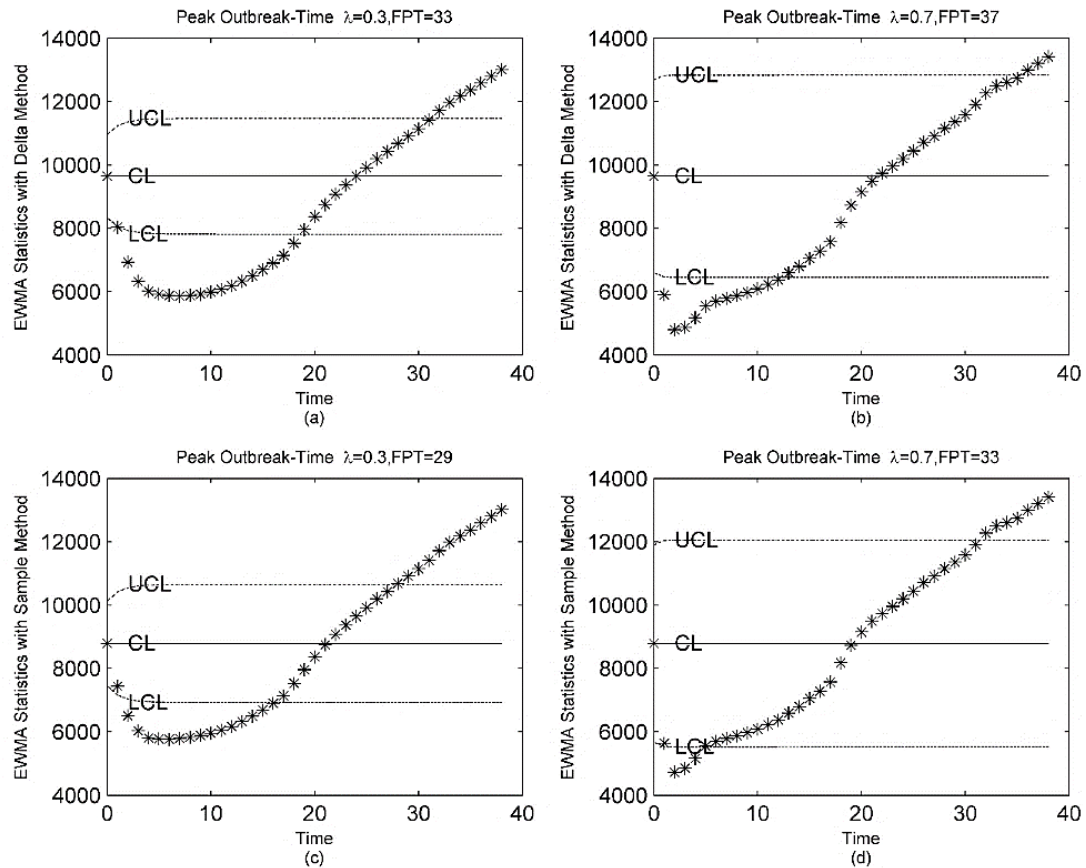
Figure 12. The EWMA control chart performances for the before peak outbreak stage of the Samut Sakhon COVID-19 outbreak according to distance using (a,b) the delta method and (c,d) the sample method
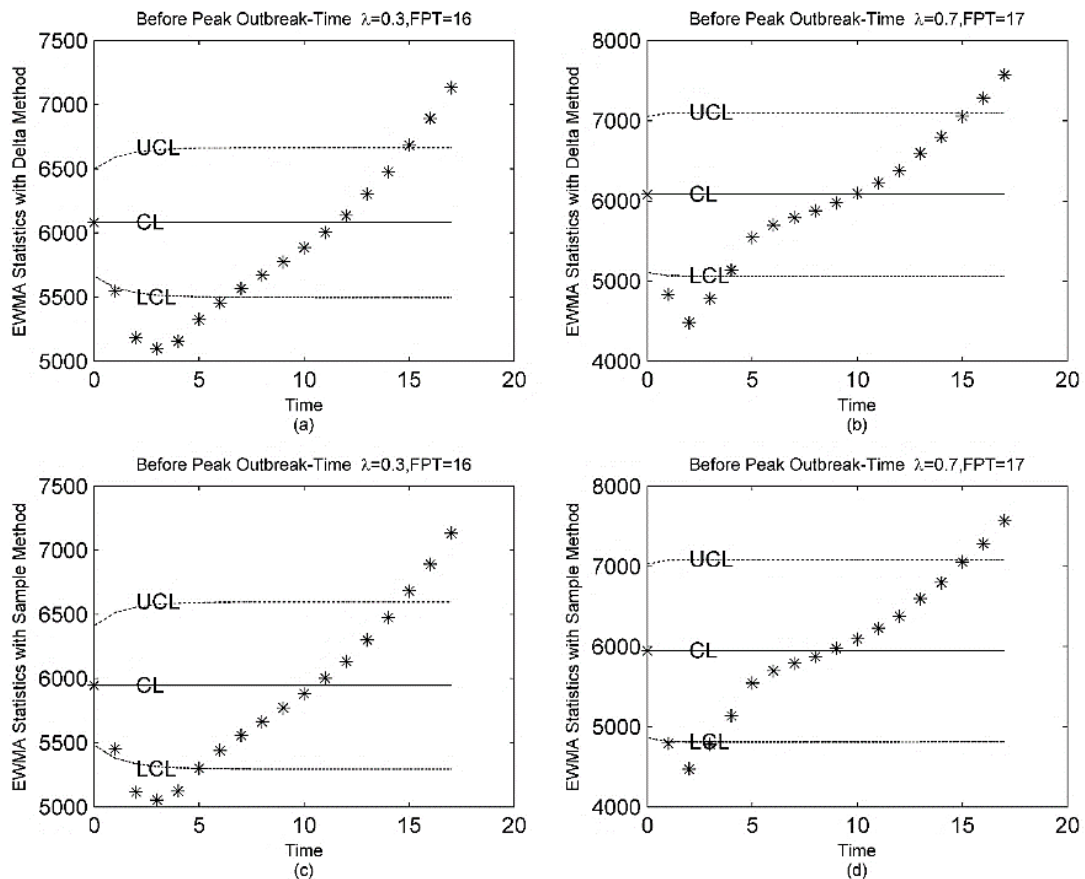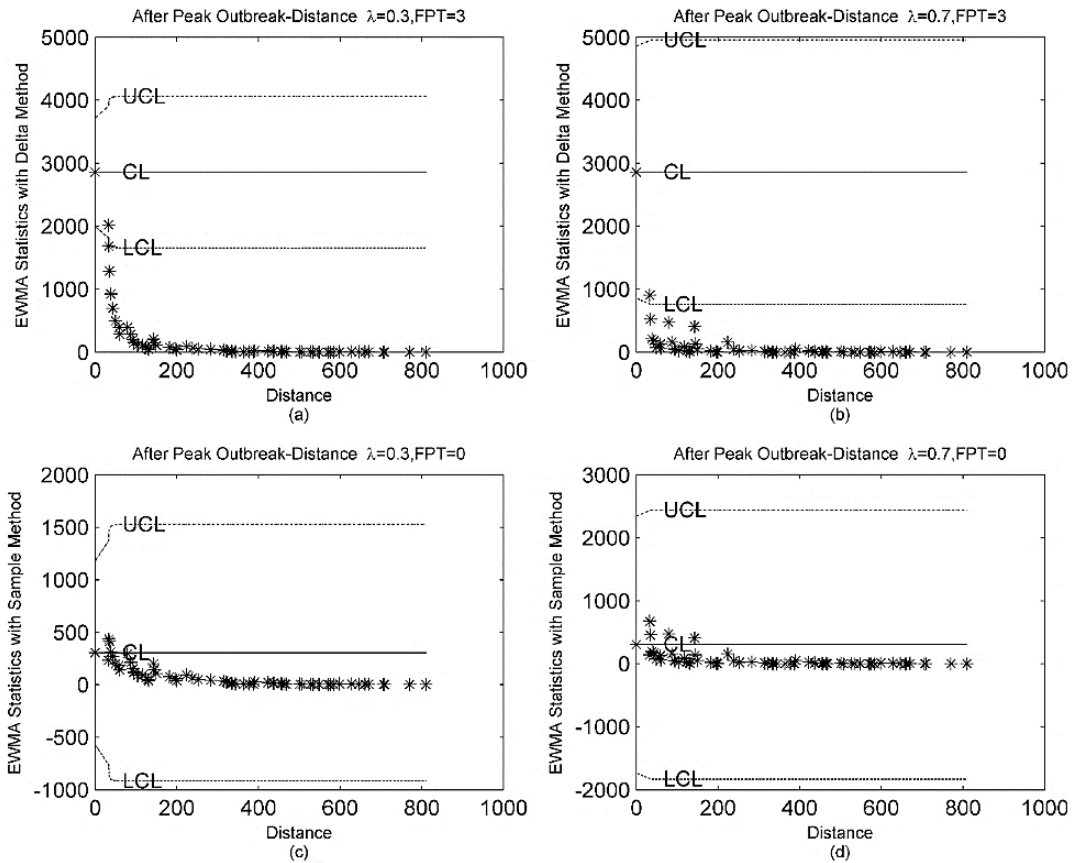

Figure 13. The EWMA control chart performances for the before peak outbreak stage of the Samut Sakhon COVID-19 outbreak according to time using (a,b) the delta method and (c,d) the sample method
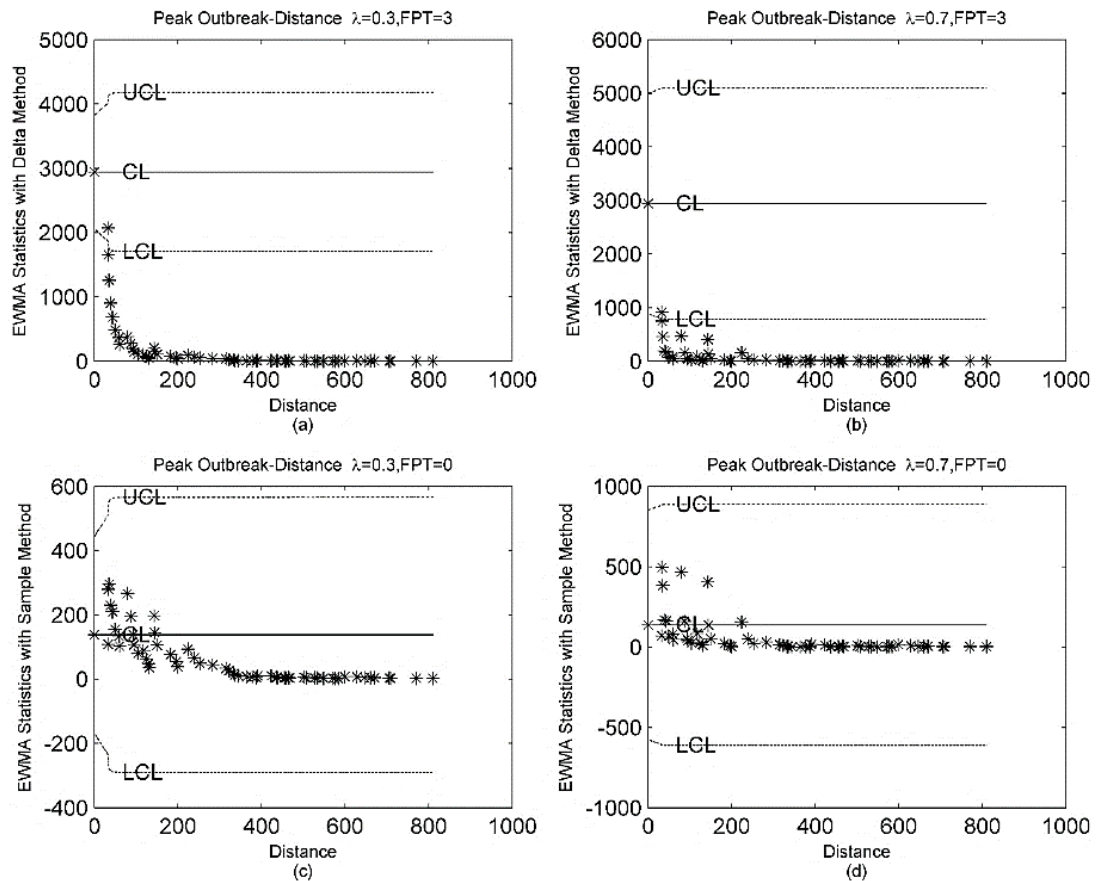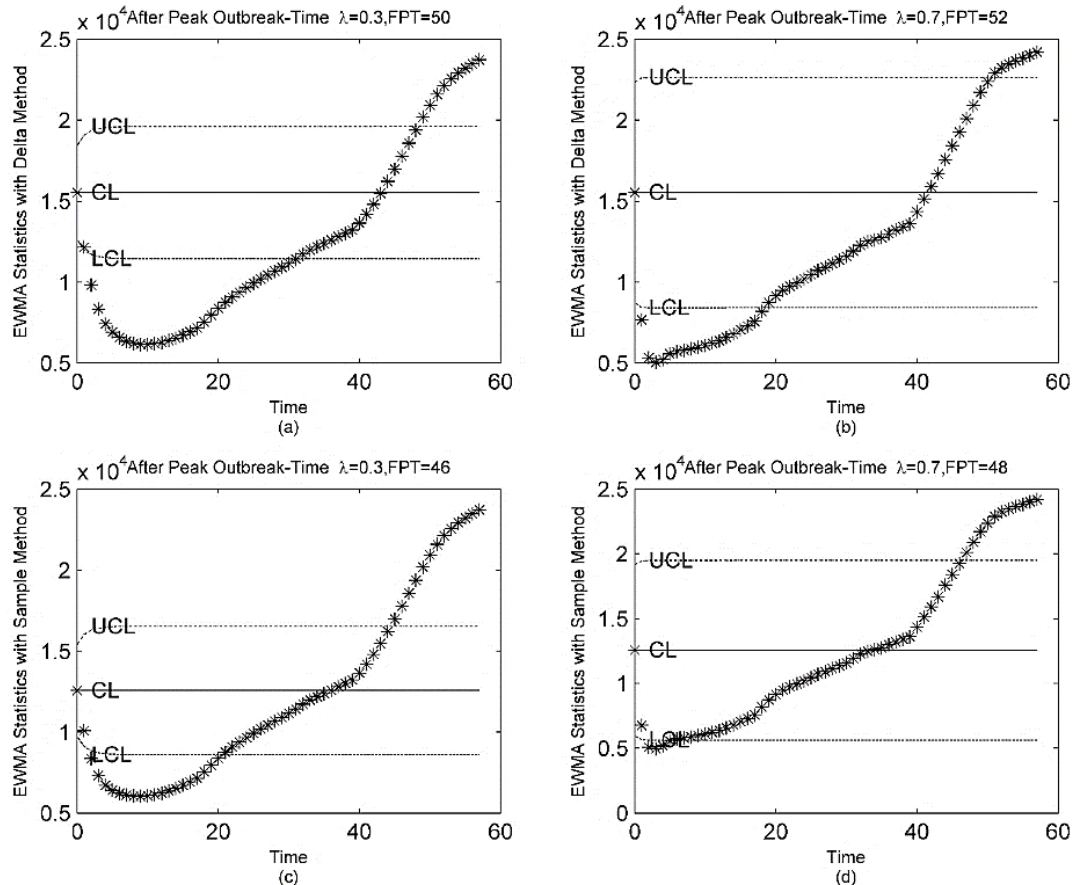
Figure 14. The EWMA control chart performances for the peak outbreak stage of the Samut Sakhon COVID-19 outbreak according to distance using (a,b) the delta method and (c,d) the sample method

Figure 15. The EWMA control chart performances for the peak outbreak stage of the Samut Sakhon COVID-19 outbreak according to time using (a,b) the delta method and (c,d) the sample method

Figure 16. The EWMA control chart performances for the after peak outbreak stage of the Samut Sakhon COVID-19 outbreak according to distance using (a,b) the delta method and (c,d) the sample method

Figure 17. The EWMA control chart performances for the after peak outbreak stage of the Samut Sakhon COVID-19 outbreak according to time using (a,b) the delta method and (c,d) the sample method

When monitoring the Samut Sakhon COVID-19 outbreak by using the EWMA control chart, in the before outbreak stage when λ= 0.3 or 0.7, an FPT of 3 was only detected by using the delta method based on distance (Figures 12-a and 12-b) whereas FPTs of 16 and 17 were detected by using the delta and sample methods, respectively, based on time (Figure 13). In the peak outbreak stage, an FPT of 3 was only detected by using the delta method based on distance when λ= 0.3 or 0.7 (Figures 14-a and 14-b, respectively). On the other hand, based on time when λ= 0.3 or 0.7, FPTs of 33 and 37 were detected by using the delta method (Figures 15-a and 15-b, respectively) and 29 and 33 by using the sample method (Figures 15-c and 15-d, respectively). In the after peak outbreak stage, an FPT of 3 when λ= 0.3 or 0.7 was only detected by using the delta method based on distance (Figures 16-a and 16-b, respectively). Meanwhile, based on time when λ= 0.3 or 0.7, FPTs of 50 and 52 were detected by using the delta method (Figures 17-a and 17-b, respectively) and 46 and 48 by using the sample method (Figures 17-c and 17-d, respectively).

The comparisons of the results between this research and the research in the literature are discussed. The results of research in the literature review are from the system of differential equations [18-26] with the time independent variable for analyzing and representing the COVID-19 outbreak; however, the results from this research are based on the single differential equation with the distance independent variable and derived it to the probabilistic analysis such as the PDF, CDF, conditional probability. The probabilistic analysis is useful for uncertainty situations, which correspond to the COVID-19 outbreak with uncertainty factors.

4- Discussion

In the literature, much of the research has been on estimating the number of COVID-19 using time as the independent variable. In the present research, the spread of COVID-19 was studied by using distance as the independent variable along with probabilistic analysis of the distance from a COVID-19 outbreak. Simulated and real data of the total number of COVID-19 cases according to the distance from the outbreak in Samut Sakhon province to other provinces along with three outbreak stages (before peak, peak, and after peak) were used in the analysis.

The PDF, CDF, the probability and conditional probability between two distances, QF, SF, HF, the median, the expected value, variance, moment, skewness, and kurtosis of the distance from the COVID-19 outbreak were derived using probability theory. The findings reveal that the estimated growth rates of the total number of COVID-19 cases according to the distance from the Samut Sakhon COVID-19 outbreak for all three outbreak stages when using the forecasting model with distance as the independent variable were negative. This means that at the beginning of the outbreak, people located at a greater distance from the center of the outbreak were less likely to become infected than those located farther away. The dispersion according to the distance from the Samut Sakhon COVID-19 outbreak was analyzed for skewness and kurtosis, which were 2.00 and 84.00, respectively, for all three outbreak stages. This means that the PDFs of the distance from the COVID-19 outbreak for all three stages were positively skewed and were thus leptokurtic (heavy-tailed) functions. The properties of the statistics in the PDFs for all three stages of the outbreak based on the distance from the Samut Sakhon COVID-19 outbreak indicated a negative growth rate (Table 2), which is in accordance with the statistics based on the simulated data (Table 1). Moreover, other functions such as the CDF, QF, SF, and HF according to the distance from the Samut Sakhon COVID-19 outbreak for all three outbreak stages also indicated a negative growth rate (Table 1).

The probability and conditional probability between two distances were evaluated to consider the probability of the COVID-19 outbreak spreading based on distance for all three outbreak stages. For all three stages of the outbreak, when the difference between two distances was large, the probability was high (Figures 11-a to 11-c), and when the difference between two distances was small, the conditional probability was high (Figures 11-d to 11-f). The probability between two distances was calculated by using the area under the PDF curve without the previous outbreak condition, whereas the conditional probability was evaluated with the previous outbreak condition. This implies that the conditional probability should be used to evaluate the probability of the outbreak spreading because it represents spreading from the outbreak center. The probability and conditional probability for the before peak and peak outbreak stages were higher and lower than those of the after peak outbreak, respectively.

Application of the EWMA control chart with the distance or time domain for detecting FPTs by using the delta or sample method was proposed for estimating the expected value and variance. For the distance domain, the delta method performed better than the sample method. Furthermore, the delta method detected the FPT at 3 for both values of the smoothing parameter and for all three outbreak stages studied (Figures 12-a and 12-b, Figures 14-a and 14-b, and Figures 16-a and 16-b), whereas the sample method detected it much later (Figures 12-c and 12-d, Figures 14-c and 14-d, and Figures 16-c and 16-d). In general, the sample method performed better than the delta method for the time domain. Although the FPTs were 16 and 17 with the delta and sample methods, respectively, in the before peak outbreak stage (Figure 13), they were 33 and 37 with the delta method (Figures15-a and 15-b) and 29 and 33 with the sample method (Figures 15-c and 15-d) for the peak outbreak stage and 50 and 52 with the delta method (Figures 17-a and 17-b) and 46 and 48 with the sample method (Figures 17-c and 17-d) for the after peak outbreak stage when λ = 0.3 or 0.7, respectively. Moreover, the FPT increased when the smoothing parameter value was increased. The findings indicate that the delta method for the EWMA control chart with distance as the independent variable is the most suitable for monitoring the Samut Sakhon COVID-19 outbreak due to detecting the lowest FTP (3). The sample method with time as the independent

variable for the EWMA control chart performed reasonably well but still attained quite a large FTP (16 was the smallest). As an illustration, an FPT of 3 for the distance domain is equivalent to that from Samut Sakhon to Bangkok (33.69 km), whereas an FPT of 16 using the time domain is equivalent to January 1, 2021, at which time the first COVID-19 cases were found in Nakhon Pathom (38.62 km), Chachoengsao (87.46 km), Surin (377.3 km), Ang Thong (117.52 km), and Ubon Ratchathani (527.74 km). In other words, the analysis of the COVID-19 outbreak based on distance as the independent variable can detect the out-of-control signal on the EWMA chart much more quickly than with time as the independent variable.

However, it is important to note that the proposed forecasting model can effectively forecast the number of COVID-19 cases when there is only one outbreak, for which probability analysis is an effective approach. However, when several outbreaks emerge nationally at around the same time, probability analysis cannot be used effectively because it is not possible to identify the outbreak area in which people were infected or the number of people infected in a given outbreak area. To mitigate this, many outbreaks can be treated as a single outbreak that covers all of Thailand for practical purposes. That is to say, Thailand would become the new center of an outbreak at the regional level.

5- Conclusion

A model was developed to analyze a COVID-19 outbreak in which the rate of spread is dependent on the distance from the center of the outbreak. The important statistics and functions for probabilistic analysis, such as the moment, expected value, variance, PDF, conditional probability, CDF, QF, skewness, kurtosis, SF, and HF, were derived and analyzed. These functions and statistics are useful for advanced statistical analysis, such as inference statistics and Bayesian statistics. Moreover, real data from a COVID outbreak that occurred at the Central Shrimp Market in Samut Sakhon province, Thailand was utilized for this research. Three outbreak stages (before peak, peak, and after peak) were used in a case study.

The findings reveal that the important statistics and functions for probabilistic analysis with real COVID-19 data corresponded with those with simulated COVID-19 data. Negative growth rates were forthcoming with the real data, based on the distance from the Samut Sakhon COVID-19 outbreak. The conditional probability is a suitable approach for analyzing the probability of COVID-19 spreading due to the outbreak based on distance because the current outbreak center is extended by the previous outbreak center. Moreover, the probability and conditional probability of the distance in the before peak and peak outbreak stages were greater than in the after peak outbreak stage. The estimated growth rate for the total number of COVID-19 cases according to the distance from the Central Shrimp Market in Samut Sakhon province meets the maximum negative value at the peak outbreak. The expected value and median meet the maximum value at the peak outbreak. It was found that, for the outbreak center identified at the beginning of the outbreak, the distance of the outbreak is inversely correlated with the probability of COVID-19 infection in both the short and long periods of the outbreak. The EWMA control chart with the delta method for estimating the expected value and variance was more suitable for monitoring a COVID-19 outbreak based on distance for all three outbreak stages than using it with the sample method based on time. In a comparison of using the distance as the independent variable rather than time, the EWMA control chart with the delta method more quickly detected the out-of-control signal for the FPT of the COVID-10 outbreak based on distance than using the sample method based on time. For further research, another application for this research is survival analysis for the COVID-19 outbreak by using the sojourn function based on this research.

6- Declarations

6-1- Author Contributions

Conceptualization, Y.A. and R.S.; methodology, R.S.; software, R.S.; validation, Y.A.; formal analysis, Y.A.; investigation, R.S.; resources, R.S.; data curation, Y.A.; writing—original draft preparation, R.S.; writing—review and editing, Y.A.; visualization, R.S.; supervision, Y.A.; project administration, Y.A.; funding acquisition, R.S. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6-3- Funding

6-4- Acknowledgements

6-5- Institutional Review Board Statement

Not applicable.

6-6- Informed Consent Statement

Not applicable.

6-7- Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Areepong, Y., & Sunthornwat, R. (2020). Predictive Models for Cumulative Confirmed COVID-19 Cases by Day in Southeast Asia. Computer Modeling in Engineering & Sciences, 125(3), 927–942. doi:10.32604/cmes.2020.012323.

[2] Wang, P., Zheng, X., Li, J., & Zhu, B. (2020). Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. Chaos, Solitons and Fractals, 139(110058). doi:10.1016/j.chaos.2020.110058.

[3] Sunthornwat, R., & Areepong, Y. (2021). Predictive Model for the Total Daily Covid-19 Cases with Herd Immunity Policy. Chiang Mai University Journal of Natural Sciences, 20(1). doi:10.12982/CMUJNS.2021.016.

[4] Brunner, N., & Kühleitner, M. (2021). Bertalanffy-Pütter models for the first wave of the COVID-19 outbreak. Infectious Disease Modelling, 6(2021), 532–544. doi:10.1016/j.idm.2021.03.003.

[5] El Aferni, A., Guettari, M., & Tajouri, T. (2021). Mathematical model of Boltzmann's sigmoidal equation applicable to the spreading of the coronavirus (Covid-19) waves. Environmental Science and Pollution Research, 28(30), 40400–40408. doi:10.1007/s11356-020-11188-y.

[6] Niazkar, M., Eryılmaz Türkkan, G., Niazkar, H. R., & Türkkan, Y. A. (2020). Assessment of Three Mathematical Prediction Models for Forecasting the COVID-19 Outbreak in Iran and Turkey. Computational and Mathematical Methods in Medicine, 2020, 1–13. doi:10.1155/2020/7056285.

[7] Khan, N., Arshad, A., Azam, M., Al-marshadi, A. H., & Aslam, M. (2022). Modeling and forecasting the total number of cases and deaths due to pandemic. Journal of Medical Virology, 94(4), 1592–1605. doi:10.1002/jmv.27506.

[8] Khan, M. A., Khan, R., Algarni, F., Kumar, I., Choudhary, A., & Srivastava, A. (2022). Performance evaluation of regression models for COVID-19: A statistical and predictive perspective. Ain Shams Engineering Journal, 13(2), 101574. doi:10.1016/j.asej.2021.08.016.

[9] Ogundokun, R. O., Lukman, A. F., Kibria, G. B. M., Awotunde, J. B., & Aladeitan, B. B. (2020). Predictive modelling of COVID-19 confirmed cases in Nigeria. Infectious Disease Modelling, 5, 543–548. doi:10.1016/j.idm.2020.08.003.

[10] Joshi, H., Azarudheen, S., Nagaraja, M. S., & Chandraketu, S. (2022). On the quick estimation of probability of recovery from COVID-19 during first wave of epidemic in India: a logistic regression approach. Statistics in Transition New Series, 23(2), 197–208. doi:10.2478/stattrans-2022-0024.

[11] Bhangu, K. S., Sandhu, J. K., & Sapra, L. (2022). Time series analysis of COVID-19 cases. World Journal of Engineering, 19(1), 40–48. doi:10.1108/WJE-09-2020-0431.

[12] Doornik, J. A., Castle, J. L., & Hendry, D. F. (2021). Modeling and forecasting the COVID-19 pandemic time-series data. Social Science Quarterly, 102(5), 2070–2087. doi:10.1111/ssqu.13008.

[13] Wang, Y., Yan, Z., Wang, D., Yang, M., Li, Z., Gong, X., Wu, D., Zhai, L., Zhang, W., & Wang, Y. (2022). Prediction and analysis of COVID-19 daily new cases and cumulative cases: times series forecasting and machine learning models. BMC Infectious Diseases, 22(1), 495. doi:10.1186/s12879-022-07472-6.

[14] Shinde, G. R., Kalamkar, A. B., Mahalle, P. N., Dey, N., Chaki, J., & Hassanien, A. E. (2020). Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art. SN Computer Science, 1(4), 1-15. doi:10.1007/s42979-020-00209-9.

[15] Ünlü, R., & Namli, E. (2020). Machine learning and classical forecasting methods based decision support systems for covid-19. Computers, Materials and Continua, 64(3), 1383–1399. doi:10.32604/cmc.2020.011335.

[16] Kuvvetli, Y., Deveci, M., Paksoy, T., & Garg, H. (2021). A predictive analytics model for COVID-19 pandemic using artificial neural networks. Decision Analytics Journal, 1, 100007. doi:10.1016/j.dajour.2021.100007.

[17] Lounis, M., Torrealba-Rodriguez, O., & Conde-Gutiérrez, R. A. (2021). Predictive models for COVID-19 cases, deaths and recoveries in Algeria. Results in Physics, 30, 104845. doi:10.1016/j.rinp.2021.104845.

[18] Wang, F., Cao, L., & Song, X. (2022). Mathematical modeling of mutated COVID-19 transmission with quarantine, isolation and vaccination. Mathematical Biosciences and Engineering, 19(8), 8035–8056. doi:10.3934/mbe.2022376.

[19] Daniel Deborah, O. (2020). Mathematical Model for the Transmission of Covid-19 with Nonlinear Forces of Infection and the Need for Prevention Measure in Nigeria. Journal of Infectious Diseases and Epidemiology, 6(5), 158. doi:10.23937/2474-3658/1510158.

[20] Ghosh, A., Roy, S., Mondal, H., Biswas, S., & Bose, R. (2022). Mathematical modelling for decision making of lockdown during COVID-19. Applied Intelligence, 52(1), 699–715. doi:10.1007/s10489-021-02463-7.

[21] Oke, A. S., Bada, O. I., Rasaq, G., & Adodo, V. (2022). Mathematical analysis of the dynamics of COVID-19 in Africa under the influence of asymptomatic cases and re-infection. Mathematical Methods in the Applied Sciences, 45(1), 137–149. doi:10.1002/mma.7769.

[22] De La Sen, M., & Ibeas, A. (2020). On a Sir Epidemic Model for the COVID-19 Pandemic and the Logistic Equation. Discrete Dynamics in Nature and Society, 2020, 1-7. doi:10.1155/2020/1382870.

[23] Zhao, H., Merchant, N. N., McNulty, A., Radcliff, T. A., Cote, M. J., Fischer, R. S. B., Sang, H., & Ory, M. G. (2021). COVID-19: Short term prediction model using daily incidence data. PLOS ONE, 16(4), e0250110. doi:10.1371/journal.pone.0250110.

[24] Riyapan, P., Shuaib, S. E., & Intarasit, A. (2021). A mathematical model of COVID-19 pandemic: A case study of Bangkok, Thailand. Computational and Mathematical Methods in Medicine, 2021, 1-11. doi:10.1155/2021/6664483.

[25] Šušteršič, T., Blagojević, A., Cvetković, D., Cvetković, A., Lorencin, I., Šegota, S. B., Milovanović, D., Baskić, D., Car, Z., & Filipović, N. (2021). Epidemiological Predictive Modeling of COVID-19 Infection: Development, Testing, and Implementation on the Population of the Benelux Union. Frontiers in Public Health, 9. doi:10.3389/fpubh.2021.727274.

[26] Campillo-Funollet, E., Van Yperen, J., Allman, P., Bell, M., Beresford, W., Clay, J., Dorey, M., Evans, G., Gilchrist, K., Memon, A., Pannu, G., Walkley, R., Watson, M., & Madzvamuse, A. (2021). Predicting and forecasting the impact of local outbreaks of COVID-19: Use of SEIR-D quantitative epidemiological modelling for healthcare demand and capacity. International Journal of Epidemiology, 50(4), 1103–1113. doi:10.1093/ije/dyab106.

[27] Khoshnaw, S. H. A., & Mohammed, A. S. (2022). Computational simulations of the effects of social distancing interventions on the COVID-19 pandemic. AIMS Bioengineering, 9(3), 239–251. doi:10.3934/bioeng.2022016.

[28] Nguyen, D. Q., Vo, N. Q., Nguyen, T. T., Nguyen-An, K., Nguyen, Q. H., Tran, D. N., & Quan, T. T. (2022). BeCaked: An Explainable Artificial Intelligence Model for COVID-19 Forecasting. Scientific Reports, 12(1), 7969. doi:10.1038/s41598-022-11693-9.

[29] Sunthornwat, R., & Areepong, Y. (2021). Predictive models for the number of cumulative cases for spreading coronavirus disease 2019 in the world. Engineering and Applied Science Research, 48(4), 432–445. doi:10.14456/easr.2021.46.

[30] Worldometer. (2021). COVID-19 Coronavirus Pandemic. Available Online: https://www.worldometers.info/coronavirus/#countries. (accessed on January 2023).

[31] DistanceFromTo (2021). Distance Between Cities and Places. Available Online: https://www.distancefromto.net. (accessed on January 2023).

[32] Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. Quarterly of Applied Mathematics, 2(2), 164–168. doi:10.1090/qam/10666.

[33] Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. Journal of the Society for Industrial and Applied Mathematics, 11(2), 431–441. doi:10.1137/0111030.

[34] Phanthuna, P., & Areepong, Y. (2021). Analytical solutions of arl for SAR(P)L model on a modified ewma chart. Mathematics and Statistics, 9(5), 685–696. doi:10.13189/ms.2021.090508.

[35] Benaroya, H., Mi Han, S., & Nagurka, M. (2005). Probability Models in Engineering and Science. Taylor & Francis, Milton Park, United Kingdom. doi:10.4324/9781003002314.

[36] Montgomery, D. C. (2020). Introduction to statistical quality control. John Wiley & Sons, Hoboken, United States.

[37] Berberian, S. K. (2012). A first course in real analysis. Springer Science & Business Media, New York, United States. doi:10.1007/978-1-4419-8548-4.