

RESEARCH ARTICLE

Estimating COVID-19 cases in Makkah region of Saudi Arabia: Space-time ARIMA modeling

Fuad A. Awwad¹, Moataz A. Mohamoud², Mohamed R. Abonazel^{1,2*}

1 Department of Quantitative Analysis, King Saud University, Riyadh, Saudi Arabia, **2** Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt

* mabonazel@cu.edu.eg

Abstract

The novel coronavirus COVID-19 is spreading across the globe. By 30 Sep 2020, the World Health Organization (WHO) announced that the number of cases worldwide had reached 34 million with more than one million deaths. The Kingdom of Saudi Arabia (KSA) registered the first case of COVID-19 on 2 Mar 2020. Since then, the number of infections has been increasing gradually on a daily basis. On 20 Sep 2020, the KSA reported 334,605 cases, with 319,154 recoveries and 4,768 deaths. The KSA has taken several measures to control the spread of COVID-19, especially during the Umrah and Hajj events of 1441, including stopping Umrah and performing this year's Hajj in reduced numbers from within the Kingdom, and imposing a curfew on the cities of the Kingdom from 23 Mar to 28 May 2020. In this article, two statistical models were used to measure the impact of the curfew on the spread of COVID-19 in KSA. The two models are Autoregressive Integrated Moving Average (ARIMA) model and Spatial Time-Autoregressive Integrated Moving Average (STARIMA) model. We used the data obtained from 31 May to 11 October 2020 to assess the model of STARIMA for the COVID-19 confirmation cases in (Makkah, Jeddah, and Taif) in KSA. The results show that STARIMA models are more reliable in forecasting future epidemics of COVID-19 than ARIMA models. We demonstrated the preference of STARIMA models over ARIMA models during the period in which the curfew was lifted.

OPEN ACCESS

Citation: Awwad FA, Mohamoud MA, Abonazel MR (2021) Estimating COVID-19 cases in Makkah region of Saudi Arabia: Space-time ARIMA modeling. PLoS ONE 16(4): e0250149. <https://doi.org/10.1371/journal.pone.0250149>

Editor: Qiang Zeng, South China University of Technology, CHINA

Received: December 27, 2020

Accepted: April 1, 2021

Published: April 20, 2021

Copyright: © 2021 Awwad et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data used in this study are third party data from <https://covid19.moh.gov.sa/> and can be accessed following the protocol outlined in the Methods section.

Funding: The Deanship of Scientific Research at King Saud University for its funding this research group (RG-1435-088).

Competing interests: The authors have declared that no competing interests exist.

Introduction

In univariate time series, we observe autocorrelation between the successive observations over a time frame. In building a time series model, the Autoregressive Moving Average (ARMA) models for spatial data can be viewed in a time series frame called space-time models. Space-time models explain the dependencies across space in the systematic dependency systems among the observations at various locations. Applications of spatial statistics cover many areas such that geostatistics, sociology, economics, and environmental.

Pfeifer and Deutsch [1] proposed the space-time autoregressive moving average (STARMA) model for some processes that can be modeled by STARMA. Models are characterized by a single random variable observed at N fixed points in space. The dependencies between the N time series are incorporated in the model through hierarchical $N \times N$ weighted matrices.

The STARMA model characterizes the spatial-temporal reliance between diverse districts by inferring the interaction between adjoining locales inside the system. So this model is more suitable for modeling geographic spatial-temporal data [1]; because this model completely considers the autocorrelation existing within the geographic spatial-temporal data set. In any case, the STARMA has certain restrictions in a few viewpoints, such as spatial and temporal non-stationary time series data modeling.

This paper is organized as follows: Section 2 reviews some literature about STARIMA Models, whereas Section 3 presents the used framework of time series data models. Section 4 introduces STARIMA models (STARMA, STAR, and STMA). Section 5 presents our empirical study on Makkah region states. Finally, Section 6 presents the conclusion and remarks.

Literature review

The STARMA was first introduced by [2]. In 1979 Pfeifer started studying "Spatial-dynamic modeling" on his Ph. D. dissertation. Later, Deutsch joined him and became his co-author on "Stationarity and invertibility regions for low-order STARMA models". Pfeifer and Deutsch successively developed this model in the eighties and published articles as follows:

In 1980a, Pfeifer and Deutsch presented the three stages of building STARIMA models: identification, estimation, and diagnostic checking, and illustrated it with a substantive example. In the same year, Pfeifer and Deutsch [3] studied the characteristics of the STAR (1,1) and STMA (1,1) models as well as the behavior and patterns of the theoretical space-time autocorrelation, partial autocorrelation functions, and initial estimation. Pfeifer and Deutsch [4] also suggested a simulation of the experiment results to compare the accuracy of three alternative estimation techniques (maximum likelihood estimation, least squares estimation, and conditional maximum likelihood estimation) of STARMA model. Pfeifer and Deutsch [5] used the likelihood ratio test to develop an independence test of the STARIMA models' residual.

Furthermore, Pfeifer and Deutsch [6] discussed the case of contemporaneously correlated variables in which the error covariance matrix (G) of the observations is non-spherical. In the same year, Pfeifer and Deutsch [7] presented Seasonal STARIMA models and discussed three stages for building the models. They illustrated the report with an example of their application. Pfeifer and Deutsch [8] also presented STARMA models where the covariance matrix of the innovations ($G = \sigma^2 I$) is not spherical. Also incorporated in these modeling procedures were tests of hypotheses concerning the form of G . The entire procedure was illustrated with two applications. Also, Pfeifer and Deutsch [9] developed approximate variances of the sample space-time autocorrelation function when the underlying process is white noise, which is needed to test the significance of the observed autocorrelations.

Recently, the STARMA models have been used in several studies, such as [10–12]. Moreover, Bayesian spatial and temporal approaches have been proposed in several studies, to take into account the effect of unobserved heterogeneity and spatio-temporal correlation, such as [13–16]

Applied previous studies for COVID-19

Perone [17] used ARIMA model to forecast the epidemic trend for COVID-19 over the period after 4 April 2020, by using the Italian epidemiological data at national and regional levels. Results suggest that COVID-19 epidemic in Italy reach the plateau, in term of cumulative cases in May 2020 and the final epidemic size in Italy maybe around 200 thousand cases.

Tandon et al [18] developed a model for forecasting future COVID-19 cases in India. In their study, confirmed, recovered and death cases of COVID-19 infection were collected in India from 22 Jan 2020 to 13 April 2020. The study indicates an ascending trend for the cases

in the coming days. The time series analysis also presents an exponential increase in the number of cases. The results for a measure of model accuracy for ARIMA, Linear Trend, Quadratic Linear, S-Curve Trend, Moving Average, Single Exponential and Double Exponential model suggests that ARIMA (2, 2, 2) model is the most accurate of all the models for forecasting future incidences as it possesses the least value for all the measures.

Iran was the first Middle East country to report a coronavirus case. Tran et al [19] developed a prediction model for COVID-19 in Iran. The model utilizes SARS-CoV-2 daily data, from 20 Feb to 4 May 2020, and ARIMA was employed to forecast the trend of the pandemic spread. The ARIMA model predicts that Iran can easily exhibit an increase in the daily total confirmed cases and the total deaths, while the daily total confirmed new cases, total new deaths, and growth rate in confirmed cases/deaths become stable in the near future. This study predicts that Iran can control the SARS-CoV-2 disease soon.

COVID-19 in KSA

Alzahrani et al [20] employed ARIMA model to forecast the expected daily number of COVID-19 cases in Saudi Arabia. They used to perform four different prediction models of ARIMA models, and determine the best model ARIMA (2,1,1). The forecasting results showed that the trend in Saudi Arabia will continue growing and may reach up to 7668 new cases per day and over 127,129 cumulative daily cases by 21 May 2020, if stringent precautionary measures were not implemented to limit the spread of COVID-19. They pointed out that training the model on more data will lead to more reliable forecasting.

Elhassan and Gaafar [21] developed both the ARIMA model and Logistic growth model to study the trend and to provide short and long-term forecasting of the prevalence COVID-19 cases in Saudi Arabia. The data analyzed in their study covered the period of 2 Mar to 21 Jun 2020. ARIMA and Logistic growth models showed excellent performance in projecting the epidemic prevalence.

Among the effects of the pandemic is the suspension of the Hajj and Umrah events in the two holy cities of Mecca and Medina that were supposed to be performed by nearly two million Muslims in mid-July. Makkah is in the middle of the cities of Jeddah and Taif. Pilgrims and Umrah performers cross those cities come to Makkah to complete their religious journey. Noticeably, the spread of the COVID-19 in the Kingdom started from Mar 2020, as a result of the movement of pilgrims and residents between those cities. These movements led to the rapid spread of COVID-19 between the cities. Thus, this paper aims to study the spatial effects (neighborhood analysis) of infected cases of COVID-19 in Makkah, Jeddah, and Taif.

ARIMA models

Box and Jenkins [22] is one of the most well-known traditional methods of time series modeling. It is also called Autoregressive Integrated Moving Average (ARIMA). Vandaele [23] explains that this method is divided into several models. It can be divided into stable and unstable models. The following is a brief presentation of the models for this method.

Autoregressive (AR) models

These models are called Autoregressive models because the variable (X_t) is linear with itself in previous time periods ($X_{t-1}, X_{t-2}, \dots, X_{t-p}$). It is possible to express the current observation of the series (X_t) as a linear function in its previous observations and random error limit, denoted by the symbol (u_t). Autoregressive models can be formulated as follows:

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + u_t \quad (1)$$

where X_t are observations of a stationary time series, u_t is a white noise process with mean zero and constant variance (σ_u^2), and $\varphi_1, \varphi_2, \dots, \varphi_p$ are autoregressive parameters of the model.

Moving Average (MA) models

These models are called moving average models because the variable (X_t) is linear with random errors limit in previous and current time periods ($u_t, u_{t-1}, u_{t-2}, \dots, u_{t-q}$). According to [24], Moving Average models can be formulated as follows:

$$X_t = u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q} \quad (2)$$

where $\theta_1, \theta_2, \dots, \theta_q$ are moving average parameters of the model.

Autoregressive Moving Average (ARMA) models

These models are called Autoregressive and Moving Average models because the variable X_t is linear with itself in previous time periods ($X_{t-1}, X_{t-2}, \dots, X_{t-p}$), and the random errors limit in previous and current time periods ($u_t, u_{t-1}, u_{t-2}, \dots, u_{t-q}$). Autoregressive Moving Average models can be formulated as follows:

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q} + u_t \quad (3)$$

Autoregressive Integrated Moving Average (ARIMA) models

Some time series models are non-stationary in mean, and thus uses differencing as for their transformation. A stationary differenced time series can be modeled with autoregressive moving average, but requires an inclusion of the differenced operator. Since the order of integration of a process is the number of differenced needed to achieve stationarity, we therefore name the model-autoregressive integrated moving average (ARIMA) model. The general form of this model is ARIMA (p, d, q):

$$W_t = \varphi_1 W_{t-1} + \dots + \varphi_p W_{t-p} - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q} + u_t \quad (4)$$

where W_t is the first difference of X_t , i.e., $W_t = X_t - X_{t-1}$.

Framework of spatial time series models

After the emergence of STARIMA models, many theoretical and applied studies, which studied and evaluated different types of these models in the statistical analysis of spatial time series data were presented, This section discusses a set of those studies in a theoretical way.

The spatial time series models have been defined as: “the models that used for the study of linear relationships between the variables in two dimensions: time and space” [3, 25]. Elhorst [25] assumes that, in the space-time model, there are relationships between the data gotten from several locales under the acceptance and the systematic dependency to the proportional distance between the locales. So, the conditional average of variable $Z_i(t)$, can be modeled by i and a linear function of the past values of the subject variables that obtained from its neighbor locales.

STARMA models

These models are called the STARMA models because the variable $Z_i(t)$ can be expressed as a spatial dependence function in the previous observations $Z_i = 1, \dots, N(t-1), \dots, Z_i = 1, \dots, N(t-K)$ and the previous random error limits $\varepsilon_i = 1, \dots, N(t-1), \dots, \varepsilon_i = 1, \dots, N(t-k)$ in addition to the current random error limit $\varepsilon_i = 1, \dots, N(t)$. The order rank is determined by the number of previous

observations. The limits of the previous random error and the spatial order of the (k^{th}) autoregressive and moving average included in the model. If the current observation $z_i(t)$ relied on the number (p) of the previous observations, the number (q) of the previous random error limits, the current random error limit $\varepsilon_i(t)$, the numbers (λ_k) of the previous observations in the site (k), the number (m_k) of the previous random error limits in the site (k), then $z_i(t)$ is said to follow a Space-Time Autoregressive Moving Average (p, q, λ_k, m_k) and it can be formulate as follows:

$$Z_i(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \varphi_{kl} L^{(l)} Z_i(t - k) + \varepsilon_i(t) - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} L^{(l)} \varepsilon_i(t - k) \tag{5}$$

where $Z_i(t)$ is the observation of the random variable at site (i) $i = 1, 2, \dots, N$, and time (t) $t = 1, 2, \dots, K$. While p is the autoregressive order, q is the moving average order, λ_k is the spatial order of the k^{th} autoregressive term, m_k is the spatial order of the k^{th} moving average term, φ_{kl} is the autoregressive parameter at temporal lag k and spatial lag l , θ_{kl} is the moving average parameter at temporal lag k and spatial lag l , $L^{(l)}$ is the $N \times 1$ vector of weights for spatial, and $\varepsilon_i(t)$ is the random normally distributed error vector at time (t) with $E[\varepsilon_i(t)] = 0$ and

$$E[\varepsilon_i(t)\varepsilon_j(t + s)] = \begin{cases} \sigma^2, & i = j, s = 0 \\ 0, & otherwise. \end{cases}$$

We refers to this model as STARMA ($p_{\lambda_1, \lambda_2, \dots, \lambda_p}, q_{m_1, m_2, \dots, m_q}$). We express the same model in vector form as

$$Z(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \varphi_{kl} W^{(l)} Z(t - k) - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} W^{(l)} \varepsilon(t - k) + \varepsilon(t), \tag{6}$$

with $E[\varepsilon(t)] = 0$ and $E[\varepsilon(t)\varepsilon(t + s)'] = \begin{cases} \sigma^2 I_{NT}, & s = 0 \\ 0, & otherwise. \end{cases}$

STAR model

Two special sub-classes of the STARMA model are considered in this study. When $q = 0$, only autoregressive terms remain, and this model is referred to as space-time autoregressive or STAR model. The STAR model is given as:

$$Z(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \varphi_{kl} W^{(l)} Z(t - k) + \varepsilon(t) \tag{7}$$

We refer to this model as STAR ($p_{\lambda_1, \lambda_2, \dots, \lambda_p}$).

STMA model

Models that contain no autoregressive terms ($p = 0$) are referred to as STMA models. STMA model is given as:

$$Z(t) = \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} W^{(l)} \varepsilon(t - k) + \varepsilon(t) \tag{8}$$

We refer to this model as STMA (q_{m_1, m_2, \dots, m_q}).

Spatial weight matrix

The specification of the form of weights $W_{ij}^{(f)}$ for various positive i 's is a matter left to the model builder who may choose weights to reflect the configuration of the physical properties of a system. For example, the county system. The $W_{ij}^{(f)}$ may be chosen to reflect the physical

properties of the observed system such as the length of the common boundary between contiguous counties i and j , the distance between the centers of counties, natural barriers such as rivers or mountains, and even the ease of accessibility of county i to county j . This last factor includes such things like the number of roads between i and j , the amount of public transportation available that connects the two county, and even the flow rates upon these avenues.

Parameters estimation

After identifying the order of the model from the STARMA model family, the next step is to estimate the parameters. This section deals with the estimation of the STARMA subclass family's parameters. It should be noted that the method of estimating STARMA models are not the same for each subclass. This is because the STAR is in linear form, whereas the STMA and mixed STARMA are in nonlinear form.

Maximum likelihood estimation

The best estimates of the φ_{kl} and θ_{kl} from many points of view are the maximum likelihood estimates (MLEs) [1]. This is because our basic model formulation has errors that are purely white noise. The distribution of ε is multivariate normal with mean zero and variance-covariance matrix, $\sigma^2 I_{NT}$. We define its probability density function as follow:

$$f(\varepsilon|\Phi, \Theta, \sigma^2) = (2\pi)^{-\frac{TN}{2}} |\sigma^2 I_{NT}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} \varepsilon' \varepsilon\right) = (2\pi)^{-\frac{TN}{2}} (\sigma^2)^{-\frac{TN}{2}} \exp\left(-\frac{S(\Phi, \Theta)}{2\sigma^2}\right) \quad (9)$$

where $S(\Phi, \Theta) = \varepsilon' \varepsilon = \sum_{i=1}^N \sum_{t=1}^T \varepsilon_i(t)^2$.

Since the $\varepsilon(t)$ are the unobservable random errors and the $Z(t)$ are the quantities actually observed, it is necessary to recursively calculate the $\varepsilon(t)$ from the observed $Z(t)$. The appropriate equation is given as

$$\varepsilon(t) = Z(t) + \sum_{k=1}^p \sum_{l=0}^{l_k} \varphi_{kl} W^{(l)} Z(t-k) - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} W^{(l)} \varepsilon(t-k) \text{ for } t = 1, 2, \dots, T \quad (10)$$

The conditional likelihood function of Φ , Θ , and σ^2 is given as

$$L(\Phi, \Theta, \sigma^2|Z) = (2\pi)^{-\frac{TN}{2}} (\sigma^2)^{-\frac{TN}{2}} \exp\left(-\frac{S_*(\Phi, \Theta)}{2\sigma^2}\right) \quad (11)$$

where $S_*(\Phi, \Theta)$ is the function of conditional sum of squares: $S_*(\Phi, \Theta) = \hat{\varepsilon}' \hat{\varepsilon}$, where the $\hat{\varepsilon}$ vector is calculated via Eq (10) with $Z(t)$ and $\varepsilon(t)$ set equal to zero for $t < 1$. The conditional MLEs of Φ , Θ , σ^2 are given as $\hat{\sigma}^2 = (NT)^{-1} S_*(\hat{\Phi}, \hat{\Theta})$, and $\hat{\Phi}$, $\hat{\Theta}$ that minimize $S_*(\Phi, \Theta)$.

COVID-19 in Makkah region

The data collected relates to the total reported confirmed cases of COVID-19 of three districts of Saudi Arabia, i.e., Makkah, Jeddah, and Taif of Makkah region. The dataset was obtained from 31 May to 11 October 2020 from the application programming interface (<https://covid19.moh.gov.sa/>). We employ ARIMA and STARIMA models utilizing the neighbor spatial weights matrices.

The objective is to show how to choose the appropriate model to fit our data. It is necessary to know whether or not there are spatial effects resulting from curfew, and determine the type of spatial interaction effects to consider for the data. To this end, we will employ the three-stage model building procedures proposed by Box and Jenkins in modeling daily confirmed patient's data of the Covid-19 pandemic for both ARIMA and STARIMA.

Table 1. Descriptive statistics of COVID-19 cases.

Location	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Jeddah	3	577	140.903	139.561	1.285	3.711
Makkah	8	623	132.828	110.63	1.494	5.140
Taif	0	306	68.776	77.469	1.234	3.539

<https://doi.org/10.1371/journal.pone.0250149.t001>

Descriptive statistics of COVID-19 time series

Descriptive statistics of the prevalent confirmed cases of COVID-19 from 31 May to 11 October 2020 are given in Table 1. That is, the prevalence of COVID-19 in Jeddah, Makkah, and Taif cases as reported during this period.

Implementation

Models were implemented using R (Version 3.6). Fig 1 shows a flow chart for the implementation of STARIMA and ARIMA models in our analysis.

Preliminary stage

The preliminary analysis of the data was done by using time plots of the series as shown in Fig 2. A visual inspection of the time series plot indicates that the confirmed cases of COVID-19

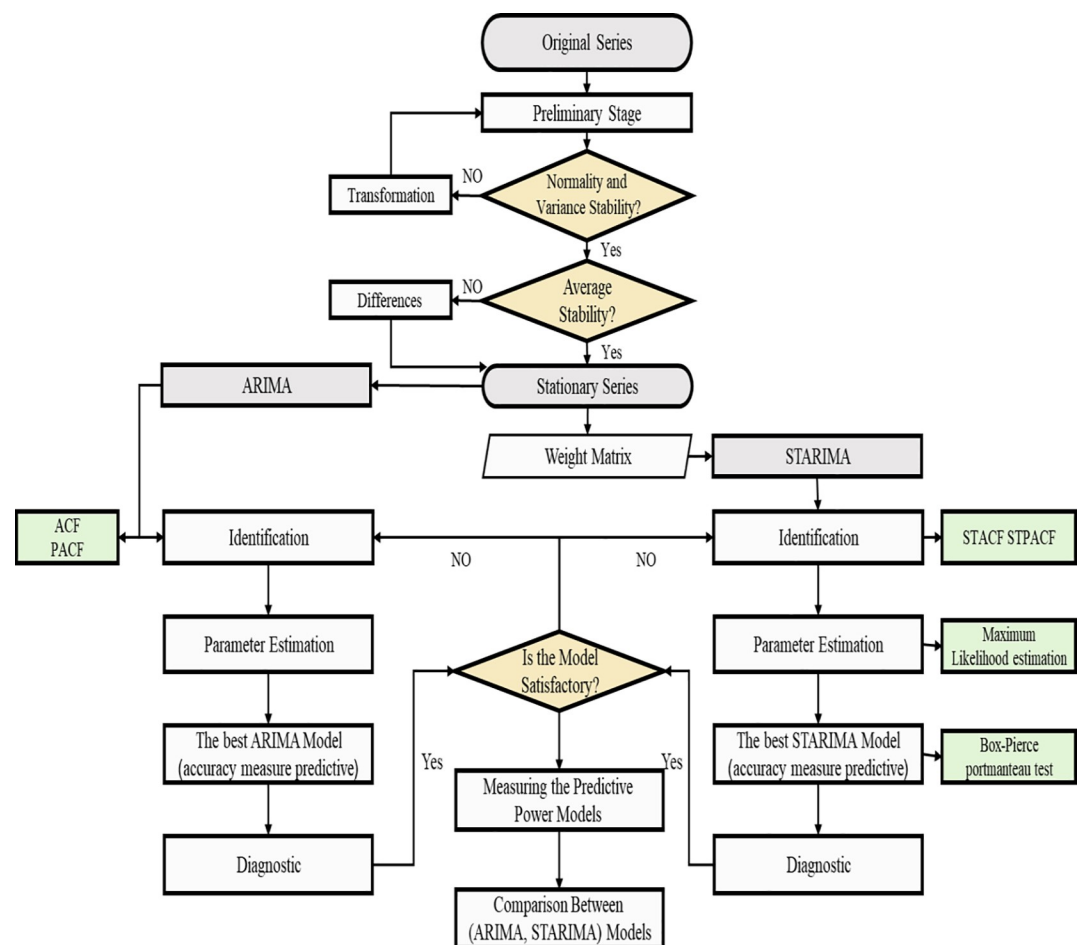


Fig 1. Applied study steps for ARIMA and STARIMA model.

<https://doi.org/10.1371/journal.pone.0250149.g001>

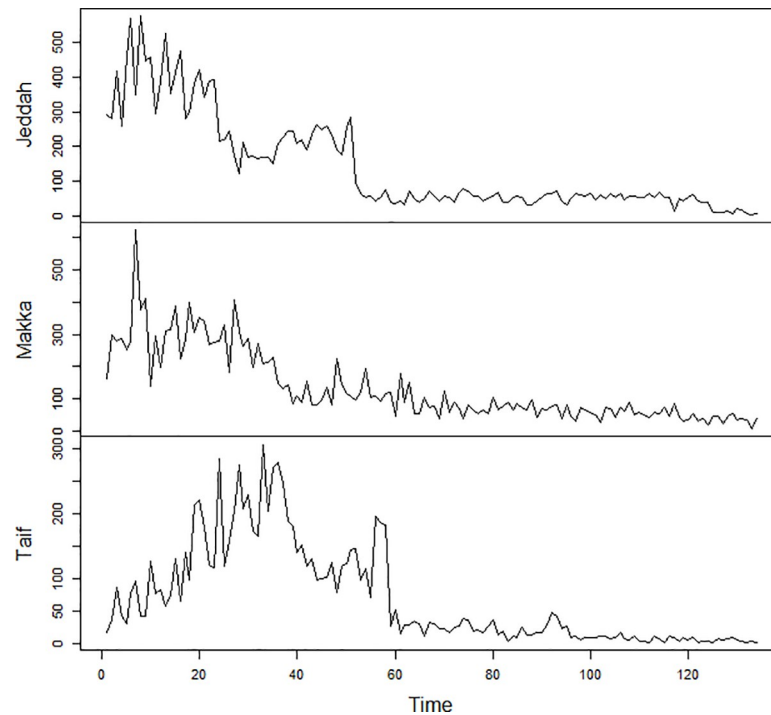


Fig 2. Time series plot of daily confirmed cases of COVID-19.

<https://doi.org/10.1371/journal.pone.0250149.g002>

in three districts of Saudi Arabia are not stationary series. So, the data must be transformed to achieve stationary. The series was transformed by taking the first difference of the natural logarithm of the values of each series to obtain stationary, the equation representing the transformation is given by: $W_t = \ln(X_t) - \ln(X_{t-1})$. After the transformation, Fig 3 shows that W_t series is stationary and have no trend. The value of $d = 1$ in ARIMA and STARIMA models.

ARIMA models for COVID-19 in Makkah region

The univariate ARIMA model is fitted for the three districts separately, according to Box-Jenkins ARIMA approach.

Identification. To identify the value of the two parameters p and q from ACF and PACF of the differenced series and compare the two plots shown in Fig 4. The ACF of all series has a significant spike at time lag (1), described by the first order of moving average. Note that the time lag (0) was ignored. The PACF of the three series have a significant spike at time lag (1) and lag (2), which is described by the order of the autoregressive process.

Parameters estimation. Base on the ACF and PACF from the first stage and "auto.arma" function in R package "forecast," the Box-Jenkins ARIMA model has been fitted for the three locations separately. As a result, it was the best models as shown in Table 2. The autoregressive and moving average parameters of our models for the three districts (Jeddah, Makkah, and Taif). After the candidate model selection and parameters estimation by the maximum likelihood estimation method. Furthermore, we can note that all moving average parameters of three models are statistically significant at 0.05.

Diagnostics. According to Box-Jenkins approach, diagnostic testing of a model involves checking the normality and the stationary of the residuals [26]. Fig 5 shows that the residuals are distributed normally. From the autocorrelation function of the residual, it is clear that all the autocorrelation coefficients are between the two limits of the confidence interval. This

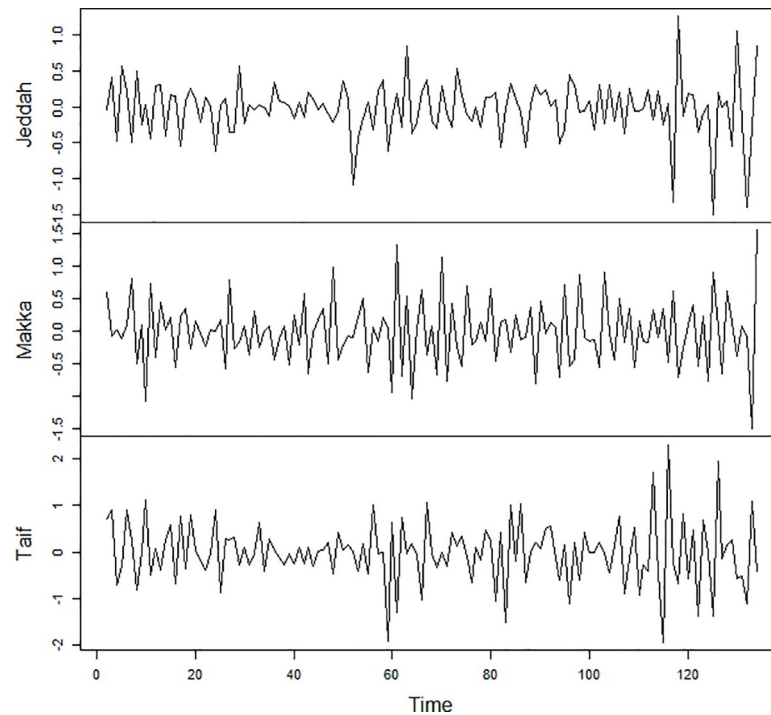


Fig 3. Time series plot of the differenced daily confirmed cases of COVID-19.

<https://doi.org/10.1371/journal.pone.0250149.g003>

indicates the independence of the residuals and the satisfaction of the assumption of independence.

STARIMA models

In this stage, we consider the multivariate STARIMA model for the three districts together.

Construction spatial weight matrix. For the STARIMA model, a spatial weight matrix was created. First order spatial neighborhood spatial matrices in Table 3 were gotten agreeing to the spatial neighborhood relationship of the three counties in Makkah region. We assumed the diagonal elements of the first-order contiguity matrix equal zero. The spatial weight matrix was also constructed by assigning equal weight to each neighbor.

STARMA model fitting

In this article, we estimated the STARMA model using the three-stage procedure according to [1]. As discussed in methodology section, STARMA estimation procedure is an extension of the Box-Jenkins methodology in spatio-temporal set up. As in ARIMA, it has three stages of model building: the model identification, estimation, and diagnostic checking.

Identification. Model identification is the most vital stride to utilize the forms of spatio-temporal models: STAR, STAMA, and STARMA. To identify the value of the parameters, we used STACF, and STPACF of the differenced series for the three locations (see Fig 6). According to Fig 6, we can conclude that the appropriate model for the data is one of the following: STARIMA (0,1,1₀), STARIMA (0,1,1₁), or STARIMA (1₀,1,1₁).

Parameters estimation. The estimated parameters of the identified models (STARIMA (0,1,1₀), STARIMA (0,1,1₁) and STARIMA (1₀,1,1₁)), using the maximum likelihood method, given in Table 4, alongside their standard errors, t.values, and p.values.

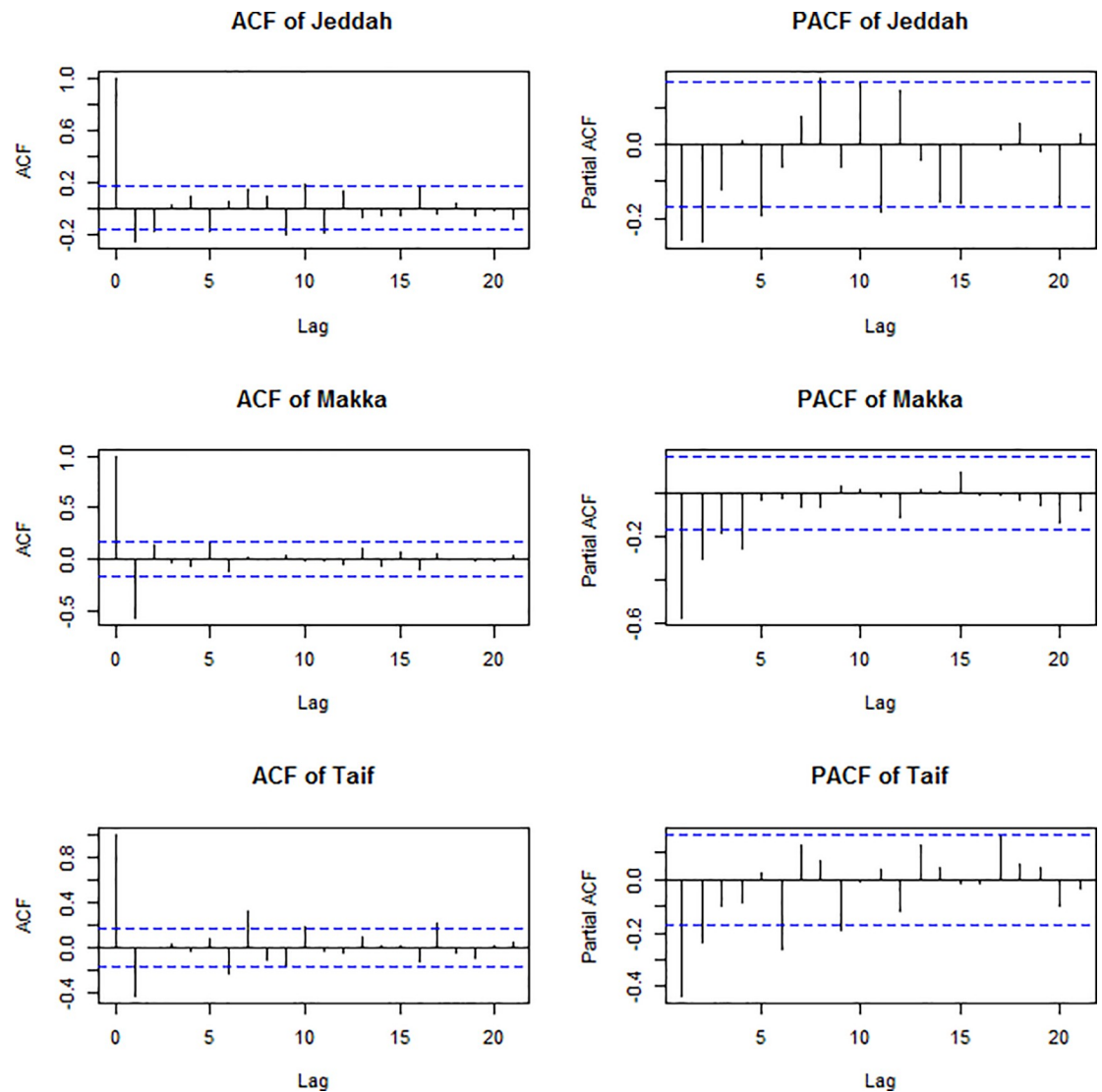


Fig 4. ACF and PACF of the differenced series.

<https://doi.org/10.1371/journal.pone.0250149.g004>

Choosing the best spatial model. To choose the best spatial model, we depended on the following accuracy measures: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), to compare the forecasting performances of the STARIMA models. According to the results in Table 5, the lowest MAE, MSE, and RMSE values is STARIMA $(1_0, 1, 1_1)$, so the best spatial model is STARIMA $(1_0, 1, 1_1)$ for the dataset.

Diagnostics. The diagnostic checking stage involves checking the independence of the residuals of STARIMA $(1_0, 1, 1_1)$ model. We applied the multivariate Box-Pierce non-correlation test (Chi-square = 51.89, df = 42, p.value = 0.141), since the p.value of Box-Pierce test more than 0.05, then we can't reject non-correlation hypothesis, this means that the residuals were uncorrelated (independent). Furthermore, from the STACF and STPACF functions of the residual (see Fig 7), it was clear that the independence of the residuals and the satisfaction of the independence assumption.

Table 2. The results of ARIMA model.

Location	Model	Statistic	AR Parameter	MA Parameter	Intercept
Jeddah	ARIMA (0,1,1)	Estimate		-0.4112***	
		SE		0.0892	
		t.value		-4.6088	
		p.value		0.0001	
Makkah	ARIMA (1,1,1)	Estimate	-0.2172	-0.7431***	-0.0177**
		SE	0.1112	0.0858	0.0067
		t.value	-1.9529	-8.6646	-2.6330
		p.value	0.0530	0.0001	0.0095
Taif	ARIMA (0,1,1)	Estimate		-0.6009***	
		SE		0.0748	
		t.value		-8.0354	
		p.value		0.0001	

Note
 , * indicate statistical significance at the 0.05, 0.001, respectively.

<https://doi.org/10.1371/journal.pone.0250149.t002>

Measuring the predictive power of models

MAE, MSE, and RMSE measures were computed to compare the forecasting performances of the best traditional ARIMA model and the best STARIMA models under considerations for all

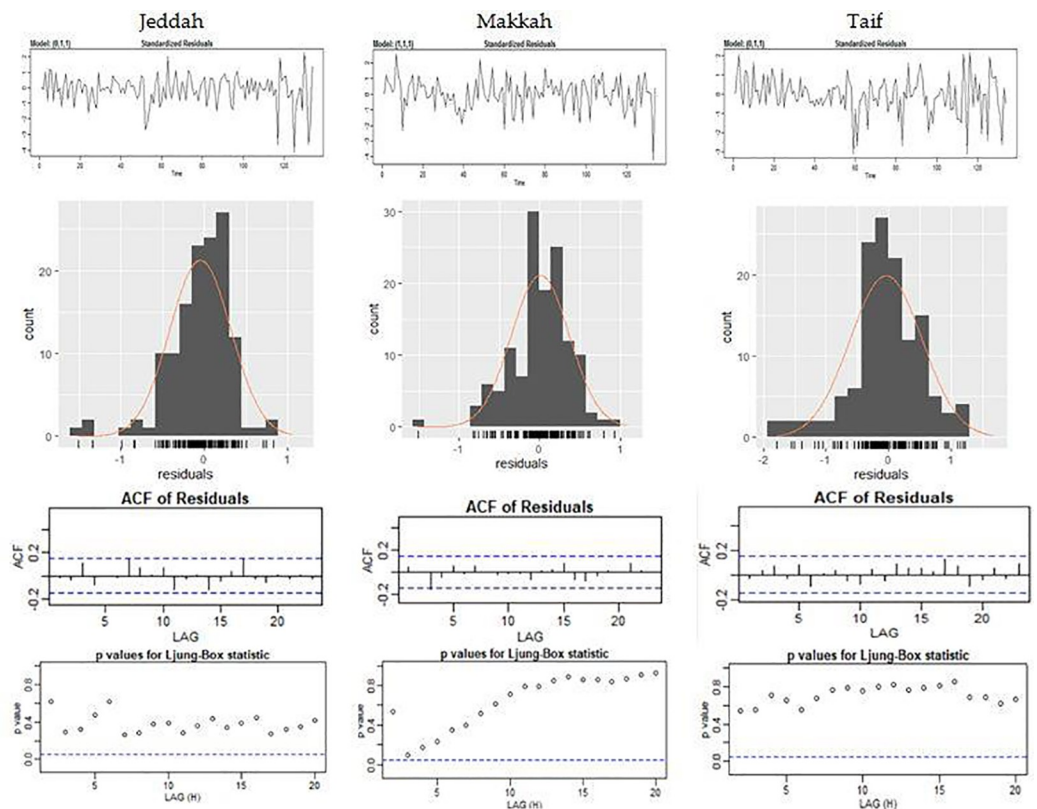


Fig 5. Residuals analysis plots of the three ARIMA model.

<https://doi.org/10.1371/journal.pone.0250149.g005>

Table 3. Spatial weight matrix of order and one.

Spatial weight matrix of order zero			
Location	Jeddah	Makkah	Taif
Jeddah	1	0	0
Makkah	0	1	0
Taif	0	0	1

First order spatial weight matrix			
Location	Jeddah	Makkah	Taif
Jeddah	0	1	0
Makkah	0.5	0	0.5
Taif	0	1	0

<https://doi.org/10.1371/journal.pone.0250149.t003>

the locations separately after the curfew. According to the results in Table 6, the lowest MAE, MSE, and RMSE values of the proposed STARIMA model for Jeddah, Makkah, and Taif after the curfew, it is confirmed that the interpretation of the curfew STARIMA model performed better than the Box-Jenkins ARIMA model in the locations for training dataset.

Conclusion

Box-Jenkins univariate ARIMA models are most popularly used in univariate time series data, whereas their applications are limited when it comes to multivariate spatio-temporal time series data. Unlike the univariate ARIMA models, STARIMA models have a smaller number of parameters. For the illustrated dataset, the STARIMA model has only three parameters for the three locations, whereas ARIMA model has numerous parameters, under such cases over parameterization may prompt to lower sum of squares of residuals. In conclusion, we studied the trend pattern of the COVID-19 outbreak in Makkah, Jeddah, and Taif in Saudi Arabia. We found out that the best prediction model for forecasting the trend of daily confirmed cases in Makkah, Jeddah, and Taif is STARIMA $(1_0, 1, 1_1)$.

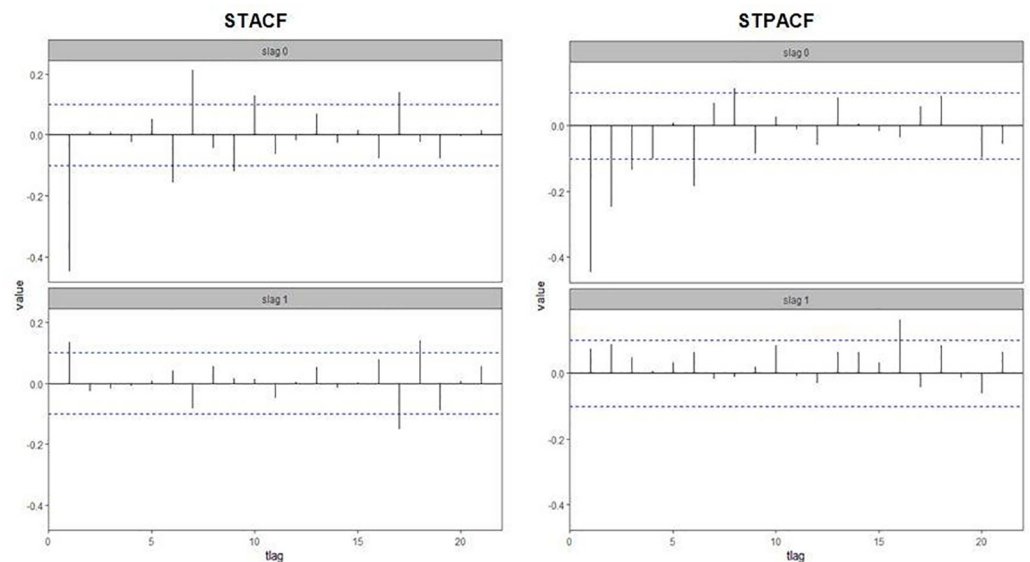


Fig 6. STACF and STPACF of the differenced series.

<https://doi.org/10.1371/journal.pone.0250149.g006>

Table 4. The results of the estimated STARIMA models.

Spatial Order	Model	Statistic	Spatial		
			AR Parameters	MA Parameters	
			ϕ_{10}	θ_{10}	θ_{11}
Lag(0)	STARIMA (0,1,1 ₀)	Estimate		-0.5882***	
		SE		0.0508	
		t.value		-11.5840	
		p.value		< 2.2e-16	
Lag(1)	STARIMA (0,1,1 ₁)	Estimate		-0.5807***	0.1976**
		SE		0.05132	0.0611
		t.value		-11.3159	3.2334
		p.value		< 2.2e-16	0.0013
	STARIMA (1 ₀ ,1,1 ₁)	Estimate	-0.0991	-0.48063***	0.1967**
		SE	0.0866	0.1013	0.0627
		t.value	-1.1443	-4.7431	3.1353
		p.value	0.2532	2.942e-06	0.0018

Note
 , * indicate statistical significance at the 0.05, 0.001, respectively.

<https://doi.org/10.1371/journal.pone.0250149.t004>

Table 5. Accuracy measures of STARIMA models.

Location	Measure	STARIMA (0,1,1 ₀)	STARIMA (0,1,1 ₁)	STARIMA (1 ₀ ,1,1 ₁)
Jeddah	MAE	0.1668	0.1710	0.1560
	MSE	0.0516	0.0544	0.0487
	RMSE	0.2273	0.2333	0.2206
Makkah	MAE	0.1800	0.1775	0.1706
	MSE	0.0557	0.0538	0.0507
	RMSE	0.2360	0.2319	0.2252
Taif	MAE	0.2489	0.2493	0.2482
	MSE	0.1167	0.1156	0.1150
	RMSE	0.3416	0.3400	0.3391

<https://doi.org/10.1371/journal.pone.0250149.t005>

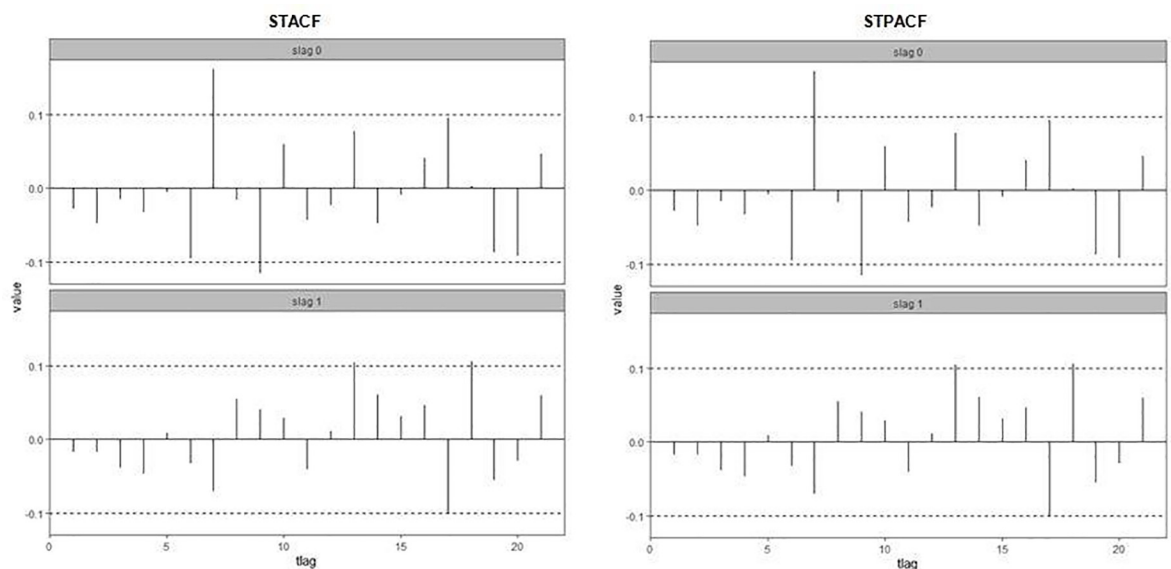


Fig 7. STACF and STPACF of STARIMA (1₀,1,1₁) residuals.

<https://doi.org/10.1371/journal.pone.0250149.g007>

Table 6. Accuracy measures of predictive power for ARIMA and STARIMA models.

Location	Models	MAE	MSE	RMSE
Jeddah	ARIMA (0,1,1)	0.2619	0.1358	0.3685
	STARMA (1 ₀ ,1,1 ₁)	0.1560	0.0487	0.2206
Makkah	ARIMA (1,1,1)	0.2712	0.1285	0.3585
	STARMA (1 ₀ ,1,1 ₁)	0.1706	0.0507	0.2252
Taif	ARIMA (0,1,1)	0.4247	0.3303	0.5747
	STARMA (1 ₀ ,1,1 ₁)	0.2482	0.1150	0.3391

<https://doi.org/10.1371/journal.pone.0250149.t006>

Based on the results obtained, one can infer that the STARIMA model outperforms the univariate ARIMA model. The outperformance of the STARIMA model over the univariate ARIMA model could be due to the inclusion of spatial information i.e. neighboring effect in the form of spatial weight matrix. Accordingly, we recommend using spatial analysis for predicting the COVID-19 confirmation cases and limiting the increase of these cases by taking appropriate preventive measures such as social distancing, curfew, banning gatherings, etc.

As future work for our study, we can use the updated spatial-temporal models, such as Bayesian spatial-temporal models [13–16] or spatial fixed/random effects panel models [27–29].

Author Contributions

Conceptualization: Moataz A. Mohamoud, Mohamed R. Abonazel.

Data curation: Moataz A. Mohamoud.

Formal analysis: Fuad A. Awwad, Moataz A. Mohamoud.

Funding acquisition: Fuad A. Awwad.

Investigation: Mohamed R. Abonazel.

Methodology: Moataz A. Mohamoud, Mohamed R. Abonazel.

Project administration: Fuad A. Awwad, Moataz A. Mohamoud, Mohamed R. Abonazel.

Resources: Fuad A. Awwad, Moataz A. Mohamoud, Mohamed R. Abonazel.

Software: Fuad A. Awwad, Mohamed R. Abonazel.

Validation: Fuad A. Awwad, Mohamed R. Abonazel.

Visualization: Mohamed R. Abonazel.

Writing – original draft: Moataz A. Mohamoud.

Writing – review & editing: Fuad A. Awwad, Mohamed R. Abonazel.

References

1. Pfeifer P.; Deutsch S. A three-stage iterative procedure for space-time modelling. *Technometrics* 1980, 22, 35–47.
2. Martin R.; Oeppen J. The identification of regional forecasting models using space time correlation functions. *Transactions of the Institute of British Geographers* 1975, 95–118.
3. Pfeifer P.; Deutsch S. Identification and Interpretation of First Order Space-Time ARMA Models. *Technometrics* 1980, 22, 397–408.
4. Pfeifer P.; Deutsch S. A comparison of estimation procedures for the parameters of the star model. *Communications in Statistics-Simulation and Computation* 1980, 9, 255–270.

5. Pfeifer P.; Deutsch S. Independence and sphericity tests for the residuals of space-time ARMA models. *Communications in Statistics-Simulation and Computation* 1980, 9, 533–549.
6. Pfeifer P.; Deutsch S. Variance of the sample space-time correlation function of contemporaneously correlated variables. *SIAM Journal on Applied Mathematics* 1981, 40, 133–136.
7. Pfeifer P.; Deutsch S. Seasonal space-time ARMA modelling. *Geographical Analysis* 1981, 13, 117–133.
8. Pfeifer P.; Deutsch S. Space-time ARMA modeling with contemporaneously correlated innovations. *Technometrics* 1981, 23, 401–409.
9. Pfeifer P.; Deutsch S. Variance of the sample space-time autocorrelation function. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1981, 43, 28–33.
10. Kurt S., & Tunay K. B. (2015). STARMA models estimation with Kalman filter: the case of regional bank deposits. *Procedia-Social and Behavioral Sciences*, 195, 2537–2547.
11. Zhao Y., Ge L., Zhou Y., Sun Z., Zheng E., Wang X., et al. (2018). A new Seasonal Difference Space-Time Autoregressive Integrated Moving Average (SD-STARIMA) model and spatiotemporal trend prediction analysis for hemorrhagic fever with renal syndrome (HFRS). *PLoS one*, 13(11), e0207518. <https://doi.org/10.1371/journal.pone.0207518> PMID: 30475830
12. Zhang X., Chen G., Wang J., Li M., & Cheng L. (2019). A GIS-Based Spatial-Temporal Autoregressive Model for Forecasting Marine Traffic Volume of a Shipping Network. *Scientific Programming*, 2019.
13. Rushworth A., Lee D., & Mitchell R. (2014). A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. *Spatial and spatio-temporal epidemiology*, 10, 29–38. <https://doi.org/10.1016/j.sste.2014.05.001> PMID: 25113589
14. Liu Y., Lund R. B., Nordone S. K., Yabsley M. J., & McMahan C. S. (2017). A Bayesian spatio-temporal model for forecasting the prevalence of antibodies to Ehrlichia species in domestic dogs within the contiguous United States. *Parasites & Vectors*, 10(1), 1–14. <https://doi.org/10.1186/s13071-017-2068-x> PMID: 28274248
15. Zeng Q., Wen H., Huang H., Abdel-Aty M., 2017. A Bayesian spatial random parameters Tobit model for analyzing crash rates on roadway segments. *Accid. Anal. Prev.* 100, 37–43. <https://doi.org/10.1016/j.aap.2016.12.023> PMID: 28088033
16. Wen H., Zhang X., Zeng Q., & Sze N. N. (2019). Bayesian spatial-temporal model for the main and interaction effects of roadway and weather characteristics on freeway crash incidence. *Accident Analysis & Prevention*, 132, 105249. <https://doi.org/10.1016/j.aap.2019.07.025> PMID: 31415995
17. Perone G. An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy. *medRxiv*, 2020.
18. Tandon H.; Ranjan P.; Chakraborty T.; Suhag V. Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future. *arXiv preprint arXiv:2004.07859*, 2020.
19. Tran T.; Pham L; Ngo Q. Forecasting epidemic spread of SARS-CoV-2 using ARIMA model (Case study: Iran). *Global Journal of Environmental Science and Management*, 6(Special Issue (Covid-19)), 2020, 1–10
20. Alzahrani S.I.; Aljamaan I.A.; Al-Fakih E.A. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *Journal of Infection and Public Health* 2020, 13, 914–919. <https://doi.org/10.1016/j.jiph.2020.06.001> PMID: 32546438
21. Elhassan T.; Gaafar A. Mathematical modeling of the COVID-19 prevalence in Saudi Arabia. *medRxiv*, 2020.
22. Box G.; Jenkins G. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.
23. Vandaele W. *Applied Time Series and Box Jenkins Models*. Academic Press, New York, 1983.
24. Bakar N.; Rosbi S. Autoregressive Integrated Moving Average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction. *International Journal of Advanced Engineering Research and Science* 2017, 4, 130–137.
25. Elhorst J. P. (2014). *Spatial econometrics: from cross-sectional data to spatial panels* (Vol. 479, p. 480). Heidelberg: Springer.
26. Abonazel M. R.; Abd-Elftah A. I. Forecasting Egyptian GDP using ARIMA Models. *Reports on Economics and Finance* 2019, 5, 35–47.
27. Kapoor M, Kelejian HH, Prucha IR. Panel Data Models with Spatially Correlated Error Components. *Journal of Econometrics* 2007; 140(1): 97–130.
28. Lee LF, Yu J. Some Recent Developments in Spatial Panel Data Models. *Regional Science and Urban Economics* 2010a; 40(5): 255–271.

29. Youssef Ahmed H., Abonazel Mohamed R., and Shalaby Ohood A. "Determinants of Per Capita Personal Income in the US: Spatial Fixed Effects Panel Data Modeling." *Journal of Advanced Research in Applied Mathematics and Statistics* 5, no. 1 & 2 (2020): 1–13.