УДК 616-036.22:616.98:578.834-07
DOI 10.59598/ME-2305-6053-2026-119-2-146-159

M. A. Sorokina[1], I. V. Korshukov[1], N. K. Omarbekova[1], D. T. Alibiyeva[1], Y. A. Drobchenko[1]*

SHORT-TERM FORECASTING USING SARIMA MODELS: AN APPLICATION TO EPIDEMIOLOGICAL TIME SERIES DATA WITH COVID-19 IN KAZAKHSTAN AS AN EXAMPLE

[1]Department of Informatics and Biostatistics, NC JSC «Karaganda Medical University» (100008, Republic of Kazakhstan, Karaganda c., Gogolya str., 40; e-mail: info@qmu.kz)

***Yelena Alexandrovna Drobchenko** – Department of Informatics and Biostatistics, NC JSC «Karaganda Medical University»; 100008, Republic of Kazakhstan, Karaganda c., Gogolya str., 40; e-mail: Drobchenko@qmu.kz

Time series forecasting is a pivotal tool across multiple disciplines, particularly in epidemiology, where precise predictions can guide resource allocation and inform public health strategies. This study explores the application of Seasonal Autoregressive Integrated Moving Average (SARIMA) models for short-term forecasting of time series data, using epidemiological data from Kazakhstan related to COVID-19 as a practical case study. The dataset encompasses total confirmed cases, daily ambulatory care patients, and daily hospitalized patients, spanning from May 30, 2022, to December 14, 2022, with forecasts extending 10 days forward to December 24, 2022. The analysis leverages SARIMA's ability to capture both seasonal and non-stationary patterns, making it highly effective for modeling complex epidemiological dynamics.

The methodology involves several key steps: data preprocessing through natural-log transformation to stabilize variance, stationarity testing using autocorrelation function (ACF) plots, and differencing to address non-stationarity. The optimal SARIMA models identified were (5,3,1)(1,1,0)_7 for total cases, (1,2,2)(0,1,1)_7 for ambulatory care patients, and (2,2,1)(2,1,1)_7 for hospitalized patients, with forecasting accuracies of 99%, 97%, and 91%, respectively. These high accuracies, measured via Mean Absolute Percentage Error (MAPE), underscore SARIMA's robustness in short-term predictions. Residual analysis, including Shapiro-Wilk and Ljung-Box tests, confirmed that the models' residuals were normally distributed and independent, validating their suitability for forecasting.

This study highlights SARIMA's versatility in capturing weekly seasonal patterns, as observed in the 7-day cycles within the COVID-19 data, which reflect reporting or behavioral trends. While the example focuses on epidemiological data, the methodology is broadly applicable to other domains, such as ecological monitoring or economic forecasting, where seasonal time series are prevalent. The findings demonstrate that SARIMA models provide a reliable framework for short-term forecasting, offering actionable insights for public health interventions and resource planning, with the COVID-19 dataset serving as an illustrative example of the approach's efficacy and adaptability.

Key words: time series forecasting; SARIMA models; short-term prediction; epidemiological data; seasonal patterns; ARIMA; Kazakhstan (as example)

INTRODUCTION

Time series analysis and forecasting are essential tools in data-driven decision-making across numerous sectors. These methods allow for the modeling of sequential data to predict future values based on historical patterns, accounting for trends, seasonality, and irregularities. While traditionally applied in fields such as finance and meteorology, time series models have proven valuable in epidemiology for anticipating disease dynamics and resource needs. This paper focuses on the general methodology of using SARIMA models for short-term forecasting, illustrated through an epidemiological example involving COVID-19 data from Kazakhstan.

The SARIMA framework extends the basic ARIMA model by incorporating seasonal components, making it suitable for data with periodic fluctuations. Historical applications include forecasting infectious disease incidences like hepatitis B [33], pertussis [24, 31], tuberculosis [34], malaria [18, 30], and hemorrhagic fever [32]. More recent studies have adapted SARIMA for emerging challenges, such as those posed by global health events [2, 3, 7, 8, 10, 11, 19, 27, 29] demonstrating its robustness for short-term predictions [11, 14, 19, 20, 27, 29]. Variations and comparisons with other models [4, 5, 9, 12, 16, 23, 28] further underscore its utility when seasonality is present.

In this context, we apply SARIMA to forecast key metrics from a real-world dataset, using COVID-19 statistics from Kazakhstan as an exemplar. As of January 5, 2023, Kazakhstan reported 1,305,649 confirmed COVID-19 cases, providing a rich time series for illustration. The dataset includes daily cumulative cases, patients under ambulatory care, and hospitalized patients, sourced from official reports [1]. However, the primary emphasis is on the forecasting technique itself – how to preprocess data, identify model parameters, evaluate residuals, and assess accuracy – rather than the specific disease. This method can be generalized to other time series, such as air quality indices in ecology or patient inflows in general healthcare planning.

Prior work in Kazakhstan has explored mathematical modeling for health risks [22, 35] and resource estimation using compartmental models like SEIR [15, 25, 26], but these often assume constant parameters or focus on long-term scenarios. Our study complements this by showcasing SARIMA's strengths in short-term, data-driven forecasting without such assumptions, using the COVID-19 period as a practical demonstration.

The aim of this study was to outline a comprehensive approach to SARIMA-based short-term forecasting, analyzing and predicting time series variables while using epidemiological data from Kazakhstan as an example. We aim to provide a blueprint for applying this method in diverse contexts, emphasizing model selection, validation, and interpretation.

MATERIALS AND METHODS

The methodology presented here is designed for general time series forecasting using SARIMA models, with an epidemiological dataset serving as an example. Data were sourced from official daily reports [1], covering total cases, ambulatory care patients, and hospitalized patients from May 30, 2022, to December 14, 2022 (training set), with December 15-24, 2022, used for validation.

SARIMA Model Overview

SARIMA extends ARIMA by adding seasonal terms, denoted as

$$(p,d,q)(P,D,Q)\_s \quad (1),$$

where:

p, q = non-seasonal autoregressive and moving average orders;

d = non-seasonal differencing order;

P, Q = seasonal autoregressive and moving average orders;

D = seasonal differencing order;

s = seasonality period (e.g., 7 for weekly cycles).

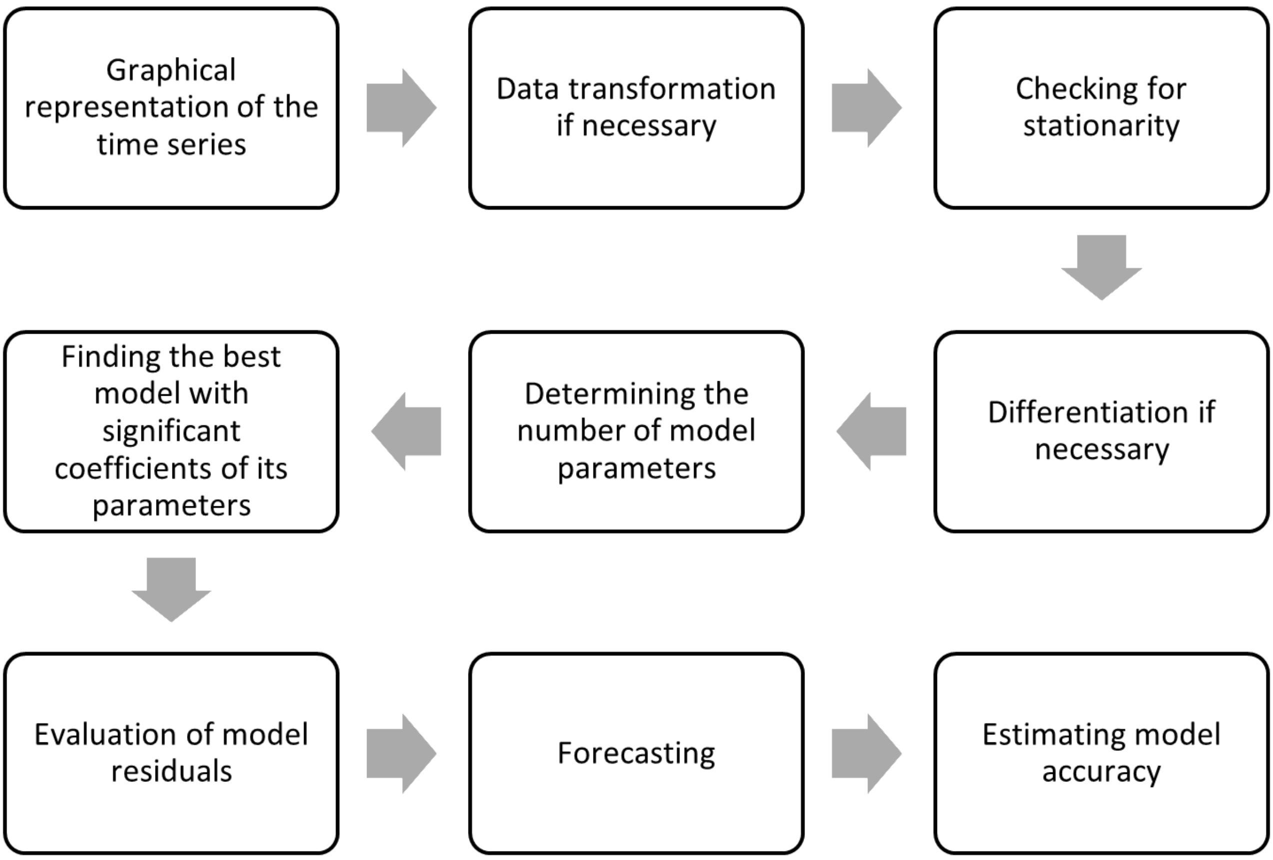

Figure 1 – Th algorithm to fit a SARIMA model

The model equation is [13]:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{P} \Phi_i y_{t-si} + e_t + \sum_{i=1}^{q} \theta_i e_{t-i} + \sum_{i=1}^{Q} \Theta_i e_{t-si} \qquad (2),$$

where:

y_t = the value of the explore variable in time ;

c = intercept;

ϕ = coefficient of each parameter ;

θ = coefficient of each parameter ;

e_t = the residuals or errors in time ;

Φ = the coefficient of each parameter ;

Θ = the coefficient of each parameter Q;

s = the seasonality of the model**.**

Algorithm for SARIMA fitting is presented in figure 1.

Graphical representation of time series

The first step was to plot time series of analyzable variables (daily total COVID-19 cases, daily number of COVID-19 patients under ambulatory care and daily number of COVID-19 hospitalized patients) for visible analysis of trend and seasonality.

Data transformation

The high variability of time series values complicates the process of creating a model for forecasting and reduces the accuracy of this model, therefore, with high data variability, it is necessary to apply a preliminary transformation of the source data. A natural-log transformation of the data was performed to stabilize this variability. In this study, the coefficient of variation (CV) was used to assess the variability of the data before and after natural-log transformation.

$$CV = \frac{s}{\bar{y}} * 100\% \qquad (3),$$

where:

CV = coefficient of variation

<u>s</u> = standard deviation

y = mean

Checking for stationarity

Most time series models require the data to be stationary. A stationary time series is a series whose behavior in the present and future coincides with the behavior in the past, i.e. properties are not affected by changing the origin of time. It is possible to determine whether a time series is stationary by looking at the form of the autocorrelation function (ACF) on their graphical plots. If the ACF does not decrease to zero or at a very slow decay, this suggests non-stationarity.

Differentiation if necessary

ARIMA (SARIMA) is an extension of ARMA models for non-stationary time series, which can be made stationary by taking differences of some order from the original time series. Box and Jenkins [6] recommended using the next differencing approach firstly, to plot the autocorrelation function of the first-difference series. Then to iterate the previous step until the ACF looks like the one of a stationary series. I (Integrated) component refers to the use of a differencing operator of raw observations, to eliminate non-stationarity.

Determining the number of model parameters

The number of times the differencing was performed to make the series stationary is the order of d and D parameters. The number of p, P, q and Q parameters can be determined by using graphical plots of the ACF and PACF. Different shapes of the autocorrelation function are possible. Usually, the autoregressive model parameters should be included in the model if shape of the autocorrelation function represents exponential decay to zero or damped oscillations. The partial autocorrelation plot identifies order of p and P parameters in this case. Presence high values at fixed intervals on the ACF plot specifies on seasonal autoregressive component. In turn of moving average model parameters the shape of the autocorrelation function represents one or more spikes, the rest is essentially zero and the orders of q and Q are identified by where autocorrelation plot becomes zero. If the exponential decay starting after a few lags, it means both autoregressive and moving average models parameters need to be included in the model.

Finding the best model with significant coefficients of its parameters

On the next step, parameter estimation is accomplished by maximizing the likelihood of the data. This technique finds such values of the parameters more precisely, their coefficients (***c, ϕ, θ, Φ, Θ***), which maximize the probability of obtaining the data that we have observed [13]. For computing the maximum likelihood value for a particular SARIMA models we used Approximate McLeod & Sales method without backcasting [6].

Evaluation of model residuals

There are two important requirements of adequate SARIMA model. First of all, the residuals (observed minus forecast values) must be normally distributed. Secondly, the residuals must be independent of each other. The normal distribution of residuals was tested by using the Shapiro – Wilk. The residuals independence was tested by plotting the autocorrelation function and applying Ljung-Box Q-test tests to the residuals of the obtained models [21].

Forecasting

In case of COVID-19, a forecast is required of what will happen in the immediate future without much regard for what will happen in the longer term. The models were built on the time interval from 2020-05-30 to 2022-12-14. The forecast was made for the next 10 days (from 2022-12-15 to 2022-12-24).

Estimating model accuracy

Forecasting accuracy is the opposite of forecasting error. If the forecast error is large, then the accuracy is small, and vice versa. Mean absolute percentage error (MAPE) was the measure of accuracy of the model fit was used in this study [13].

Table 1 – CV values of the data before and after natural logarithm transformation

Data set	Before	After
Total COVID-19 cases	2,588	0,184
Daily number of COVID-19 patients under ambulatory care	126,9	23,05
Daily number of COVID-19 hospitalized patients	117,9	21,19

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{y(t)-\hat{y}(t)}{y(t)} * 100\% \quad (3),$$

Forecast accuracy in % = 100% - *MAPE*

where:
MAPE = mean absolute percentage error
n = number of observation;
y(t)= real value;
ŷ(t) = forecast value.

The MAPE forecast error estimate is the reciprocal of the forecast accuracy. If MAPE<10%, then the forecast is made with high accuracy, 10% < MAPE< 20% – the forecast is good, 20% < MAPE< 50% – the forecast is satisfactory and MAPE> 50% – the forecast is bad [17].

Analysis was performed using Statistica 10, adaptable to other software like R or Python. This framework is versatile, applicable to any time series exhibiting seasonality, with the COVID-19 data used here to demonstrate each step.

RESULTS

The graphical presentation of the time-series datasets showed that there was a sharp increase in cumulative cases of COVID-19 in Kazakhstan (figure 2, part a) from early July to early September 2022, there was an increase of 80,000 cases in two months.

Daily number of COVID-19 patients under ambulatory care (figure 2, part b) and number of COVID-19 hospitalized patients (figure 2, part c) had a pronounced peak at the end of July and beginning of August 2022. There was a decrease in daily number of COVID-19 patients under ambulatory care and number of COVID-19 hospitalized patients since the beginning of August to the end of September. During October 2022, the time series did not change much and plateaus were observed at that time on the presented figure 2, but since November 2022, the growth of all three studied time series has begun again. It should be noted that a new rise in the time series of the daily number of COVID-19 patients under ambulatory care was recorded from the end of November 2022, while in the time series of the daily number of COVID-19 hospitalized patients it began earlier on two weeks.

At the next stage of this work, we apply data transformation in the form of a natural logarithm. CV values (table 1) confirmed high initial variability, reduced significantly post-log transformation, facilitating modeling.

In the case of daily number of COVID-19 patients under ambulatory care and number of COVID-19 hospitalized patients, the values of the coefficient of variation (126,9 and 117,9% respectively) exceed 100%, which means that the dispersion level around the mean is huge and the standard deviation is greater than the mean. Natural logarithm data transformation made it possible to reduce the variability of these data sets by more than five times.

After data transforming, we evaluated the form of the autocorrelation function (ACF) on their graphical plots to infer whether the time series are stationary or non-stationary (figure 3).

The red dotted lines in the graphs of figure 3 mark the critical interval within which the values of the autocorrelation function are considered not to differ from zero. Analyzing the graphical presentation of autocorrelation functions we could conclude all three time series data sets were non-stationary, because the plots indicated strong serial dependencies for lags of 1 to 15 with the highest value of autocorrelation for a lag of one.

Using Box and Jenkins recommendations [6] in order to make the series stationary and to remove the serial dependency, several non-seasonal differencing transformation were performed in these series and the results of transformations were presented at the figure 4.

As can be seen from figure 4 most serial dependencies have disappeared. However, the removal of the lower order serial dependencies exposed a higher order seasonality for lag of 7. There is also a clear (seasonal) dependency for a lag of 14 (it multiples of 7). This indicates a seasonal pattern in data sets. In other words, there are days when number of patients is more and when it is less. Seasonal differencing with the lag of 7 was used for removing this dependency at the next stage of this work.

Previously, it was found that the data required logarithmic transformation and two types of differencing (nonseasonal and seasonal). At the next step we determined the number of models parameters. The order of d and D parameters was taking from the number of non-seasonal and seasonal differencing transformation. The number of p, P, q and Q parameters were determined by using graphical plots of the ACF and PACF. Parameter estimation in SARIMA models were accomplished by maximizing the likelihood of the data and the following parameter coefficients for each model were obtained and presented in table 2.

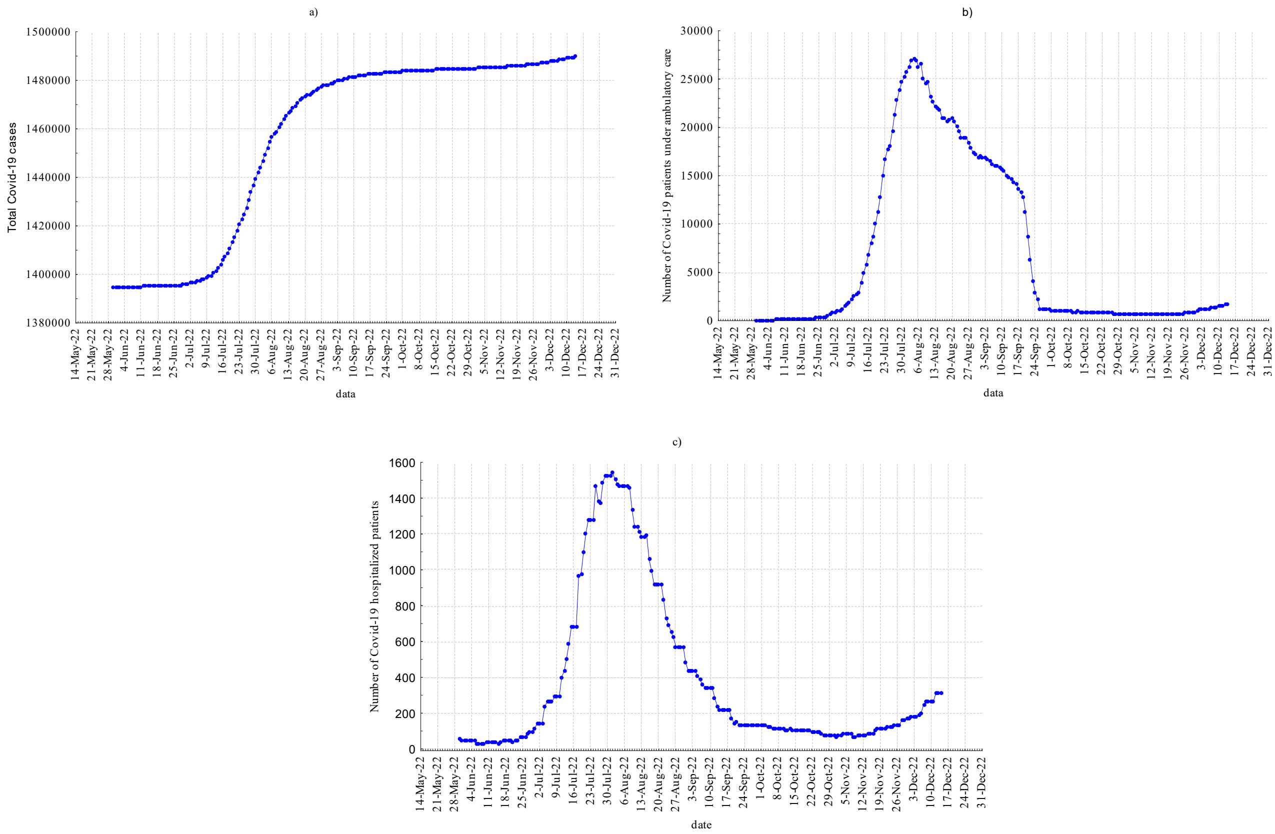

Figure 2 – Daily total COVID-19 cases (a), number of COVID-19 patients under ambulatory care (b) and number of COVID-19 hospitalized patients (c) during the period from 2020-05-30 to 2022-12-14

Table 2 – SARIMA models and their parameters

Data set/model	Coefficients		Std. Err.	t	p
Total COVID-19 cases/(5,3,1)(1,1,0)7	ϕ_1	-0,357	0,125	-2,848	0,005
	ϕ_2	-0,515	0,096	-5,359	0,000
	ϕ_3	-0,572	0,088	-6,540	0,000
	ϕ_4	-0,282	0,094	-3,012	0,003
	ϕ_5	-0,354	0,081	-4,349	0,000
	θ_1	0,354	0,122	2,906	0,004
	ϕ_1	-0,463	0,074	-6,280	0,000
Daily number of COVID-19 patients under ambulatory care/(1,2,2)(0,1,1)7	ϕ_1	-0,912	0,054	-16,82	0,000
	θ_1	-0,542	0,082	-6,644	0,000
	θ_2	0,417	0,070	5,975	0,000
	Θ_1	0,728	0,048	15,24	0,000
Daily number of COVID-19 hospitalized patients/(2,2,1)(2,1,1)7	ϕ_1	0,026	0,089	0,288	0,774
	ϕ_2	-0,246	0,082	-2,983	0,003
	θ_1	0,787	0,063	12,41	0,000
	ϕ_1	0,499	0,111	4,514	0,000
	ϕ_2	0,280	0,088	3,174	0,002
	Θ_1	0,911	0,076	12,01	0,000

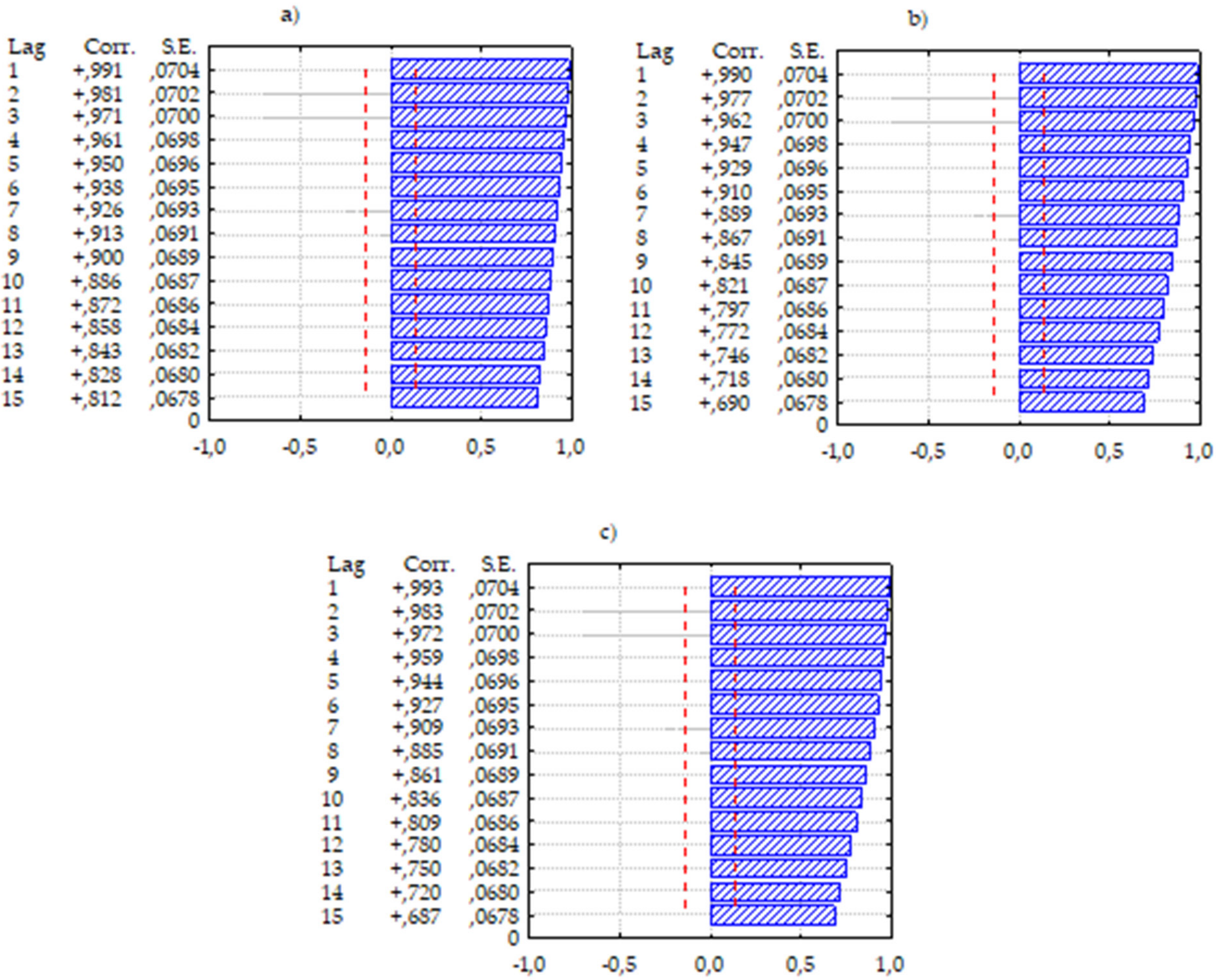

Figure 3 – Autocorrelation Function of Daily total COVID-19 cases (a), number of COVID-19 patients under ambulatory care (b) and number of COVID-19 hospitalized patients (c) for first 15 lags

Both the non-seasonal and seasonal autoregressive and moving average parameters coefficients were highly statistically significant ($p<0,05$) in all presented models (the exception was the parameter ϕ_1 in Daily number of COVID-19 hospitalized patients model, $p=0,774$).

The next step after finding the best SARIMA models was to estimate the residuals. P-values of the Shapiro – Wilk tests were greater than 0,05 consequently the residuals were subject to a normal distribution in all models (table 3).

Simultaneously there is no significant correlation in the residuals series and residuals sample autocorrelations are in (-0,2; 0,2) interval, it confirms that residuals are white noise (table 3). It is known that white noise is a process that has a constant mathematical expectation, constant variance, and zero (for all but zero lag) autocorrelation function [21]. Also, the results of Ljung-Box Q-tests (all calculated p-values greater than 0,05) for the autocorrelation pattern in the residuals indicated they were independently distributed (table 3). Featured SARIMA models respond adequately to information about the data sets because their residuals were normally distributed and independent to each other. The fulfillment of these conditions allows using the models for forecasting and the forecast of these models will be good, and the forecast intervals that are calculated considering the normal distribution will be accurate [13].

Applying SARIMA models to the training set we forecasted Total COVID-19 cases (figure 5, a), Daily number of COVID-19 patients under ambulatory care (figure 5, b) and number of COVID-19 hospitalized patients (figure 5, c) for the 10 days (from 2022-12-15 to 2022-12-24).

All forecasted values were higher than the actual, observed values in all cases. For the first five days of the forecast, the differences between the forecast and actual values were minimal. As expected, the 90% confidence intervals for the forecasted values gradually increased and reached its maximum by the tenth day of the forecast. Forecasting accuracy calculated based on mean absolute percentage error (MAPE) was presented in the table 4.

Estimating the forecasts accuracies obtained, which for all the presented models were more than

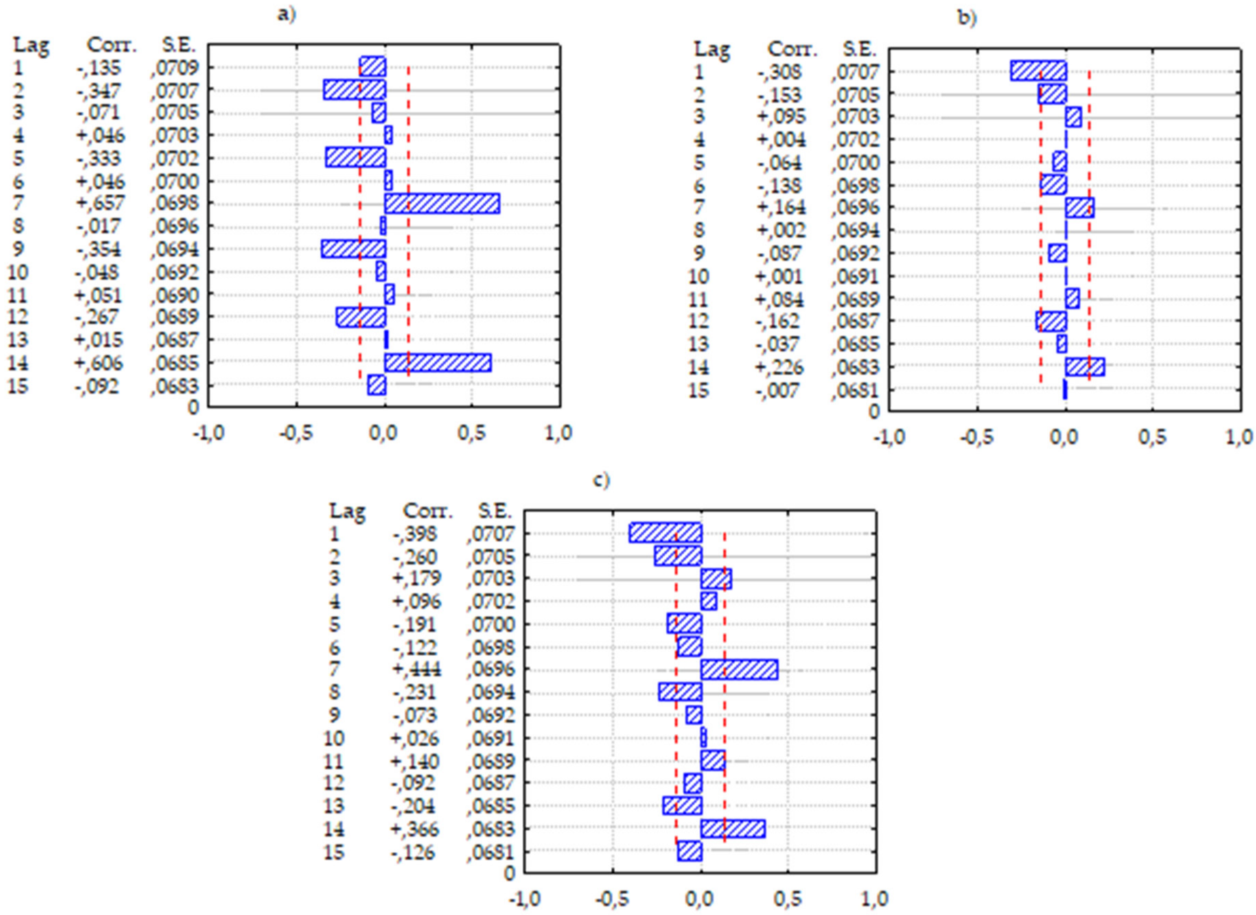

Figure 4 – Autocorrelation Function of daily total COVID-19 cases (a), number of COVID-19 patients under ambulatory care (b) and number of COVID-19 hospitalized patients (c) for first 15 lags after differencing transformation

90%, we can conclude that the forecast data were in good agreement with the actually observed data. Therefore, we can recommend these SARIMA models for a short-term forecast of total covid-19 cases, daily number of COVID-19 patients under ambulatory care and number of COVID-19 hospitalized patients.

Thereby applying the SARIMA methodology to the example dataset revealed patterns and enabled accurate short-term forecasts, illustrating the model's effectiveness. Graphical plots (Figure 2) showed trends with peaks and plateaus, indicative of seasonal influences in the time series. CV values (table 1) confirmed high initial variability, reduced significantly post-log transformation, facilitating modeling. ACF plots (figure 3) indicated non-stationarity, resolved through differencing (figure 4), exposing a 7-day seasonality. Optimal models (table 2) had significant coefficients (mostly p<0.05), confirming fit. Residuals were normally distributed (Shapiro – Wilk p>0.05) and independent (Ljung-Box p>0.05; table 3), validating the models. Forecasts (figure 5) with 90% confidence intervals aligned well with observed data, with MAPE values (table 4) indicating high accuracy (>90%). These results exemplify how SARIMA captures complex patterns in time series, providing reliable short-term predictions.

DISCUSSION

This study show cases SARIMA as a powerful tool for time series forecasting, using COVID-19 data from Kazakhstan as an example to highlight its practical application. The method's ability to handle non-stationarity and seasonality makes it broadly applicable.

Data preprocessing (log transformation, differencing) addressed variability and trends, a standard practice for improving model accuracy in diverse datasets [6, 13]. The identified 7-day cycles align with weekly patterns often seen in reporting-dependent series, as noted in similar studies [11, 19].

Model parameters (e.g., AR order indicating dependence on recent lags) provide insights into data dynamics, consistent with applications in other fields [2, 3, 7, 8, 19, 27, 29]. High accuracies (>90%) affirm SARIMA's suitability for short-term forecasting, supporting findings in public health and beyond [11, 29].

Compared to alternatives [4, 5, 9, 12, 16, 23, 28], SARIMA excels in seasonal data without requiring complex hybrids. While the example focuses on epidemiological metrics for resource planning, the approach extends to ecological monitoring (e.g., pollution levels) or economic indicators.

Table 3 – Shapiro – Wilk test, Autocorrelation Function and Box & Ljung Q test results for the residuals of SARIMA models for first 15 lags

Data set/model	Shapiro – Wilk tests	lag	Autocorr.	Std. Err.	Box & Ljung Q test	
					Q	p
Total COVID-19 cases/ SARIMA(5,3,1)(1,1,0)7	SW – W=0,983; p=0,125	1	-0,010	0,072	0,019;	0,890
		2	0,001	0,072	0,020;	0,990
		3	0,049	0,072	0,476;	0,924
		4	0,045	0,072	0,869;	0,929
		5	0,035	0,071	1,108;	0,953
		6	0,058	0,071	1,780;	0,939
		7	-0,029	0,071	1,943;	0,963
		8	0,155	0,071	6,714;	0,568
		9	-0,056	0,071	7,354;	0,600
		10	-0,153	0,070	12,08;	0,280
		11	0,086	0,070	13,59;	0,256
		12	-0,143	0,070	17,77;	0,123
		13	0,062	0,070	18,57;	0,137
		14	-0,063	0,070	19,40;	0,150
		15	-0,109	0,069	21,86;	0,112
Daily number of COVID-19 patients under ambulatory care/ SARIMA(1,2,2)(0,1,1)7	SW – W=0,984; p=0,142	1	-0,004	0,072	0,003	0,957
		2	0,012	0,072	0,030	0,985
		3	-0,009	0,072	0,047	0,997
		4	-0,023	0,071	0,153	0,997
		5	-0,040	0,071	0,467	0,993
		6	-0,126	0,071	3,593	0,732
		7	-0,074	0,071	4,696	0,697
		8	0,030	0,071	4,871	0,771
		9	-0,043	0,070	5,250	0,812
		10	-0,063	0,070	6,063	0,810
		11	-0,005	0,070	6,068	0,869
		12	-0,009	0,070	6,084	0,912
		13	-0,001	0,070	6,084	0,943
		14	0,086	0,069	7,615	0,908
		15	0,016	0,069	7,670	0,936
Daily number of COVID-19 hospitalized patients/SARIMA(2,2,1)(2,1,1)7	SW – W = 0,992; p = 0,539	1	0,002	0,072	0,000	0,982
		2	-0,036	0,072	0,249	0,883
		3	0,039	0,072	0,539	0,910
		4	-0,070	0,071	1,503	0,826
		5	-0,105	0,071	3,689	0,595
		6	0,051	0,071	4,211	0,648
		7	-0,016	0,071	4,261	0,749
		8	-0,091	0,071	5,916	0,657
		9	0,063	0,070	6,709	0,667
		10	-0,020	0,070	6,787	0,745
		11	0,059	0,070	7,488	0,758
		12	0,043	0,070	7,861	0,796
		13	-0,114	0,070	10,52	0,651
		14	-0,008	0,069	10,53	0,722
		15	0,011	0,069	10,56	0,783

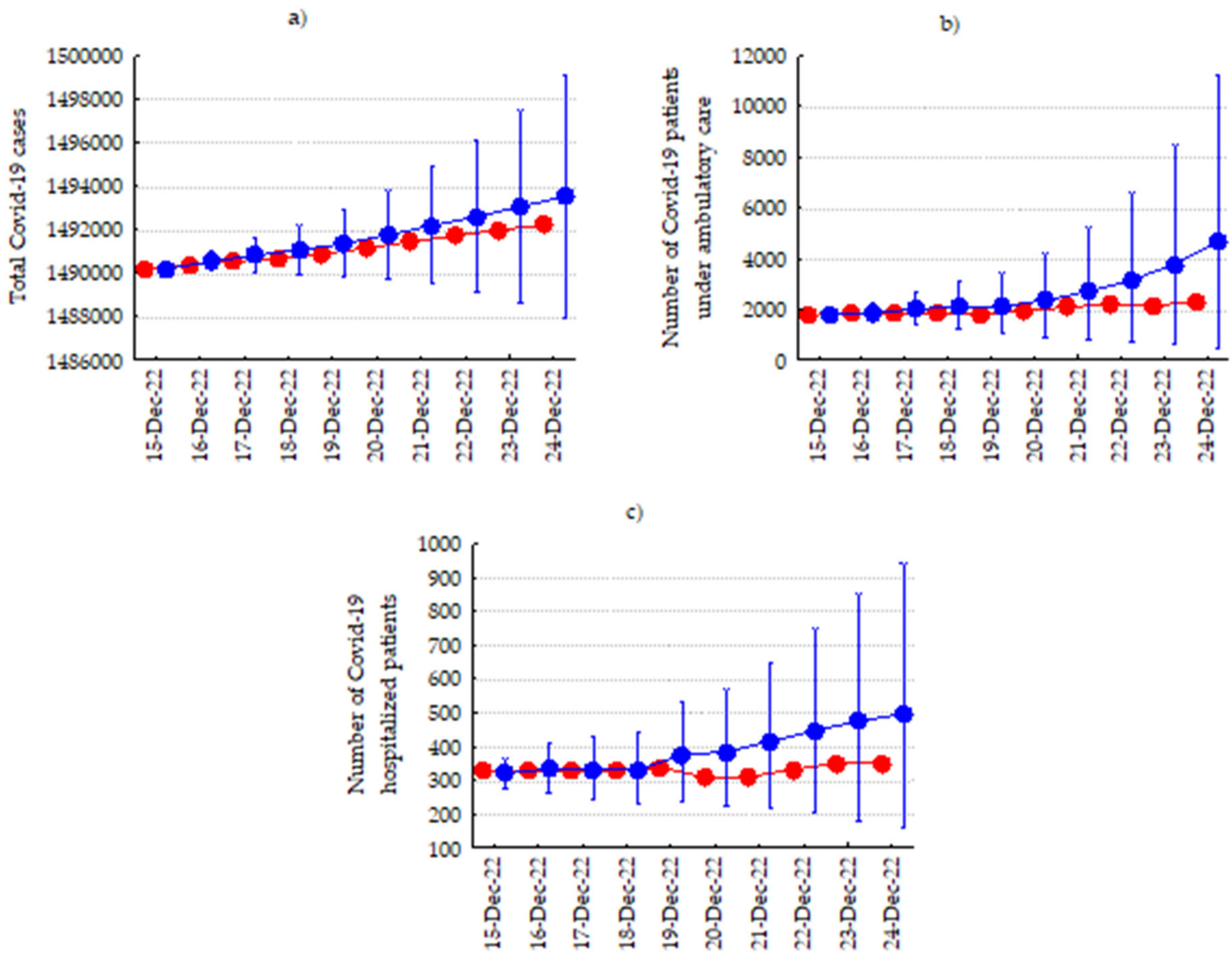

Figure 5 – Forecast of daily total COVID-19 cases (a), number of COVID-19 patients under ambulatory care (b) and number of COVID-19 hospitalized patients (c) with 90% confidence interval for the next 10 days. Observed cases – red dots, forecast cases – blue dots and lines

Table 4 – Forecasting accuracy of SARIMA models

Model	MAPE	Forecast accuracy, %
SARIMA (5,3,1)(1,1,0)7	0,202	99,8
SARIMA (1,2,2)(0,1,1)7	2,600	97,4
SARIMA (2,2,1)(2,1,1)7	9,000	91

The 10-day horizon balances accuracy and utility [11], demonstrating SARIMA's strengths over long-term models like SEIR [15, 25, 26].

Limitations. The models are optimized for short-term forecasts; longer periods may increase errors. Parameters are dataset-specific and may require retraining with new data. While illustrated with Kazakhstani data, adaptations are needed for other contexts. External factors (e.g., policy changes) not captured here could influence general applications.

CONCLUSIONS

SARIMA models offer a robust framework for short-term time series forecasting, as demonstrated using COVID-19 epidemiological data from Kazakhstan. This approach, emphasizing data transformation, stationarity, and residual validation, yields high-accuracy predictions and can be applied across domains for effective planning and resource management.

Author contributions:

M. A. Sorokina, I. V. Korshukov – study concept and design.

M. A. Sorokina, I. V. Korshukov, and N. K. Omarbekova – data collection and processing.

M. A. Sorokina – writing.

D. T. Alibiyeva, Y. A. Drobchenko – data collection and editing.

Conflict of interest:

The authors declare no conflict of interest.

REFERENCES

1. *База данных по COVID-19 Министерства здравоохранения Республики Казахстан.* https://www.gov.kz/memleket/entities/dsm?lang=ru

2. Alabdulrazzaq H., Alenezi M.N., Rawajfih Y., Alghannam B.A., Al-Hassan A.A., Al-Anzi F.S. On the accuracy of ARIMA based prediction of COVID-19 spread. *Results Phys.* 2021; 27: 104509. https://doi.org/10.1016/j.rinp.2021.104509

3. Ali M., Khan D.M., Aamir M., Khalil U., Khan Z. Forecasting COVID-19 in Pakistan. *PLoS One.* 2020; 15 (11): e0242762. https://doi.org/10.1371/journal.pone.0242762

4. Al-Turaiki I., Almutlaq F., Alrasheed H., Alballa N. Empirical Evaluation of Alternative Time-Series Models for COVID-19 Forecasting in Saudi Arabia. *Int. J. Environ. Res. Public Health.* 2021; 18 (16): 8660. https://doi.org/10.3390/ijerph18168660

5. Bhardwaj S., Alowaidi M., Bhardwaj R., Sharma S.K. Machine learned hybrid Gaussian analysis of COVID-19 pandemic in India. *Results Phys.* 2021; 30: 104630. https://doi.org/10.1016/j.rinp.2021.104630

6. Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M. *Time Series Analysis: Forecasting and Control*. New Jersey; 2016: 709.

7. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci. Total Environ.* 2020; 729: 138817. https://doi.org/10.1016/j.scitotenv.2020.138817

8. Dawoud I. Modeling Palestinian COVID-19 Cumulative Confirmed Cases: A Comparative Study. *Infect. Dis. Model.* 2020; 5: 748-754. https://doi.org/10.1016/j.idm.2020.09.001

9. De Livera A.M., Hyndman R.J., Snyder R.D. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*. 2011; 106 (496): 1513-1527. https://doi.org/10.1198/jasa.2011.tm09771

10. Ding Y., Huang R., Shao N. Time Series Forecasting of US COVID-19 Transmission. *Altern. Ther. Health Med.* 2021; 27 (S1): 4-11.

11. Duangchaemkarn K., Boonchieng W., Wiwatanadate P., Chouvatut V. SARIMA Model Forecasting Performance of the COVID-19 Daily Statistics in Thailand during the Omicron Variant Epidemic. *Healthcare (Basel).* 2022; 10 (7): 1310. https://doi.org/10.3390/healthcare10071310

12. Gecili E., Ziady A., Szczesniak R.D. Forecasting COVID-19 confirmed cases, deaths and recoveries: Revisiting established time series modeling through novel applications for the USA and Italy. *PLoS One.* 2021; 16 (1): e0244173. https://doi.org/10.1371/journal.pone.0244173

13. Hyndman R.J., Athanasopoulos G. *Forecasting: Principles and Practice.* Melbourne; 2021: 138.

14. James A., Tripathi V. Time series data analysis and ARIMA modeling to forecast the short-term trajectory of the acceleration of fatalities in Brazil caused by the corona virus (COVID-19). *Peer J.* 2021; 9: e11748. https://doi.org/10.7717/peerj.11748

15. Kabanikhin S.I., Bektemesov M.A., Bektemessov J.M. Mathematical model for medium-term COVID-19 forecasts in Kazakhstan. *Journal of Mathematics, Mechanics and Computer Science*. 2021; 3 (111): 95-106.

16. Layelmam M., Ahmed L.Y., Benchelha S.D., Yahya R. S. (). Forecasting COVID-19 in Morocco. *Journal of Clinical and Experimental Investigations*. 2020; 11: em00748. https://doi.org/10.5799/jcei/8264

17. Lewis C.D. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. Oxford; 1982: 156.

18. Malki Z., Laporta G.Z. ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound. *Neural Computing and Applications*. 2021; 7 (33): 2929-2948.

19. Mangla S., Pathak A.K., Arshad M., Haque U. Short-term forecasting of the COVID-19 outbreak in India. *Int. Health*. 2021; 13 (5): 410-420. https://doi.org/10.1093/inthealth/ihab031

20. Moffat I., Akpan E. White Noise Analysis: A Measure of Time Series Model Adequacy. *Applied Mathematics*. 2019; 10: 989-1003. https://doi.org/10.4236/am.2019.1011069.

21. Nabirova D., Taubayeva R., Maratova A., Henderson A., Nassyrova S., Kalkanbayeva M., Alaverdyan S., Smagul M., Levy S., Yesmagambetova A., Singer D. Factors Associated with an Outbreak of COVID-19 in Oilfield Workers, Kazakhstan, 2020. *Int. J. Environ. Res. Public. Health.* 2022; 19 (6): 3291. https://doi.org/10.3390/ijerph19063291

22. Perone G. Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy. *Eur. J. Health Econ.* 2022; 23 (6): 917-940. https://doi.org/10.1007/s10198-021-01347-4

23. Raycheva R., Stoilova Y., Kevorkyan A., Rangelova V. Epidemiological Prognosis of Pertussis Incidence in Bulgaria. *Folia Med. (Plovdiv).* 2020; 62 (3): 509-514. https://doi.org/10.3897/folmed.62.e49812

24. Semenova Y., Glushkova N., Pivina L., Khismetova Z., Zhunussov Y., Sandybaev M., Ivankov A. Epidemiological Characteristics and Forecast of COVID-19 Outbreak in the Republic of Kazakhstan. *J. Korean Med. Sci.* 2020; 35 (24): e227. https://doi.org/10.3346/jkms.2020.35.e227

25. Semenova Y., Pivina L., Khismetova Z., Auyezova A., Nurbakyt A., Kauysheva A., Ospanova D., Kuziyeva G., Kushkarova A., Ivankov A., Glushkova N. Anticipating the Need for Healthcare Resources Following the Escalation of the COVID-19 Outbreak in the Republic of Kazakhstan. *J. Prev. Med. Public Health.* 2020; 53 (6): 387-396. https://doi.org/10.3961/jpmph.20.395

26.Somyanonthanakul R., Warin K., Amasiri W., Mairiang K., Mingmalairak C., Panichkitkosolkul W., Silanun K., Theeramunkong T., Nitikraipot S., Suebnukarn S. Forecasting COVID-19 cases using time series modeling and association rule mining. *BMC Med. Res. Methodol.* 2022; 22 (1): 281. https://doi.org/10.1186/s12874-022-01755-x

27.Talkhi N., Akhavan Fatemi N., Ataei Z., Jabbari Nooghabi M. Modeling and forecasting number of confirmed and death caused COVID-19 in IRAN: A comparison of time series forecasting methods. *Biomed Signal. Process Control.* 2021; 66: 102494. https://doi.org/10.1016/j.bspc.2021.102494

28.Tan C.V., Singh S., Lai C.H., Zamri A.S.S.M., Dass S.C., Aris T.B., Ibrahim H.M., Gill B.S. Forecasting COVID-19 Case Trends Using SARIMA Models during the Third Wave of COVID-19 in Malaysia. *International Journal of Environmental Research and Public Health.* 2022; 19 (3): 1504. https://doi.org/10.3390/ijerph19031504

29. Wang M., Wang H., Wang J., Liu H., Lu R., Duan T., Gong X., Feng S., Liu Y., Cui Z., Li C., Ma J. A novel model for malaria prediction based on ensemble algorithms. *PLoS One.* 2019; 14 (12): e0226910. https://doi.org/10.1371/journal.pone.0226910

30. Wang M., Pan J., Li X., Li M., Liu Z., Zhao Q., Luo L., Chen H., Chen S., Jiang F., Zhang L., Wang W., Wang Y. ARIMA and ARIMA-ERNN models for prediction of pertussis incidence in mainland China from 2004 to 2021. *BMC Public Health.* 2022; 22 (1): 1447. https://doi.org/10.1186/s12889-022-13872-9

31. Wang T., Liu J., Zhou Y., Cui F., Huang Z., Wang L., Zhai S. Prevalence of hemorrhagic fever with renal syndrome in Yiyuan County, China, 2005-2014. *BMC Infect Dis*. 2016; 16: 69. https://doi.org/10.1186/s12879-016-1404-7

32. Wang Y.W., Shen Z.Z., Jiang Y. Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China. *PLoS One.* 2018; 13 (9): e0201987. https://doi.org/10.1371/journal.pone.0201987

33. Yan C.Q., Wang R.B., Liu H.C., Jiang Y., Li M.C., Yin S.P., Xiao T.Y., Wan K.L., Rang W.Q. Application of ARIMA model in predicting the incidence of tuberculosis in China from 2018 to 2019. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2019; 40 (6): 633-637. https://doi.org/10.3760/cma.j.issn.0254-6450

34. Zhussupov B., Saliev T., Sarybayeva G., Altynbekov K., Tanabayeva S., Altynbekov S., Tuleshova G., Pavalkis D., Fakhradiyev I. Analysis of COVID-19 pandemics in Kazakhstan. *J. Res. Health Sci.* 2021; 21 (2): e00512. https://doi.org/10.34172/jrhs.2021.52

TRANSLITERATION

1. *Baza dannyh po COVID-19 Ministerstva zdravoohranenija Respubliki Kazahstan*. https://www.gov.kz/memleket/entities/dsm?lang=ru

2. Alabdulrazzaq H., Alenezi M.N., Rawajfih Y., Alghannam B.A., Al-Hassan A.A., Al-Anzi F.S. On the accuracy of ARIMA based prediction of COVID-19 spread. *Results Phys.* 2021; 27: 104509. https://doi.org/10.1016/j.rinp.2021.104509

3. Ali M., Khan D.M., Aamir M., Khalil U., Khan Z. Forecasting COVID-19 in Pakistan. *PLoS One.* 2020; 15 (11): e0242762. https://doi.org/10.1371/journal.pone.0242762

4. Al-Turaiki I., Almutlaq F., Alrasheed H., Alballa N. Empirical Evaluation of Alternative Time-Series Models for COVID-19 Forecasting in Saudi Arabia. *Int. J. Environ. Res. Public Health.* 2021; 18 (16): 8660. https://doi.org/10.3390/ijerph18168660

5. Bhardwaj S., Alowaidi M., Bhardwaj R., Sharma S.K. Machine learned hybrid Gaussian analysis of COVID-19 pandemic in India. *Results Phys.* 2021; 30: 104630. https://doi.org/10.1016/j.rinp.2021.104630

6. Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M. *Time Series Analysis: Forecasting and Control*. New Jersey; 2016: 709.

7. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci. Total Environ.* 2020; 729: 138817. https://doi.org/10.1016/j.scitotenv.2020.138817

8. Dawoud I. Modeling Palestinian COVID-19 Cumulative Confirmed Cases: A Comparative Study. *Infect. Dis. Model.* 2020; 5: 748-754. https://doi.org/10.1016/j.idm.2020.09.001

9. De Livera A.M., Hyndman R.J., Snyder R.D. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*. 2011; 106 (496): 1513-1527. https://doi.org/10.1198/jasa.2011.tm09771

10. Ding Y., Huang R., Shao N. Time Series Forecasting of US COVID-19 Transmission. *Altern. Ther. Health Med.* 2021; 27 (S1): 4-11.

11. Duangchaemkarn K., Boonchieng W., Wiwatanadate P., Chouvatut V. SARIMA Model Forecasting Performance of the COVID-19 Daily Statistics in Thailand during the Omicron Variant Epidemic. *Healthcare (Basel).* 2022; 10 (7): 1310. https://doi.org/10.3390/healthcare10071310

12. Gecili E., Ziady A., Szczesniak R.D. Forecasting COVID-19 confirmed cases, deaths and recoveries: Revisiting established time series modeling through novel applications for the USA and Italy. *PLoS One.* 2021; 16 (1): e0244173. https://doi.org/10.1371/journal.pone.0244173

13. Hyndman R.J., Athanasopoulos G. *Forecasting: Principles and Practice.* Melbourne; 2021: 138.

14. James A., Tripathi V. Time series data analysis and ARIMA modeling to forecast the short-term trajectory of the acceleration of fatalities in Brazil caused by the corona virus (COVID-19). *Peer J.* 2021; 9: e11748. https://doi.org/10.7717/peerj.11748

15.Kabanikhin S.I., Bektemesov M.A., Bektemessov J.M. Mathematical model for medium-term COVID-19 forecasts in Kazakhstan. *Journal of Mathematics, Mechanics and Computer Science*. 2021; 3 (111): 95-106.

16. Layelmam M., Ahmed L.Y., Benchelha S.D., Yahya R. S. (). Forecasting COVID-19 in Morocco. *Journal of Clinical and Experimental Investigations*. 2020; 11: em00748. https://doi.org/10.5799/jcei/8264

17. Lewis C.D. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. Oxford; 1982: 156.

18. Malki Z., Laporta G.Z. ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound. *Neural Computing and Applications*. 2021; 7 (33): 2929-2948.

19. Mangla S., Pathak A.K., Arshad M., Haque U. Short-term forecasting of the COVID-19 outbreak in India. *Int. Health*. 2021; 13 (5): 410-420. https://doi.org/10.1093/inthealth/ihab031

20. Moffat I., Akpan E. White Noise Analysis: A Measure of Time Series Model Adequacy. *Applied Mathematics*. 2019; 10: 989-1003. https://doi.org/10.4236/am.2019.1011069.

21. Nabirova D., Taubayeva R., Maratova A., Henderson A., Nassyrova S., Kalkanbayeva M., Alaverdyan S., Smagul M., Levy S., Yesmagambetova A., Singer D. Factors Associated with an Outbreak of COVID-19 in Oilfield Workers, Kazakhstan, 2020. *Int. J. Environ. Res. Public. Health.* 2022; 19 (6): 3291. https://doi.org/10.3390/ijerph19063291

22. Perone G. Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy. *Eur. J. Health Econ.* 2022; 23 (6): 917-940. https://doi.org/10.1007/s10198-021-01347-4

23. Raycheva R., Stoilova Y., Kevorkyan A., Rangelova V. Epidemiological Prognosis of Pertussis Incidence in Bulgaria. *Folia Med. (Plovdiv).* 2020; 62 (3): 509-514. https://doi.org/10.3897/folmed.62.e49812

24. Semenova Y., Glushkova N., Pivina L., Khismetova Z., Zhunussov Y., Sandybaev M., Ivankov A. Epidemiological Characteristics and Forecast of COVID-19 Outbreak in the Republic of Kazakhstan. *J. Korean Med. Sci.* 2020; 35 (24): e227. https://doi.org/10.3346/jkms.2020.35.e227

25. Semenova Y., Pivina L., Khismetova Z., Auyezova A., Nurbakyt A., Kauysheva A., Ospanova D., Kuziyeva G., Kushkarova A., Ivankov A., Glushkova N. Anticipating the Need for Healthcare Resources Following the Escalation of the COVID-19 Outbreak in the Republic of Kazakhstan. *J. Prev. Med. Public Health.* 2020; 53 (6): 387-396. https://doi.org/10.3961/jpmph.20.395

26.Somyanonthanakul R., Warin K., Amasiri W., Mairiang K., Mingmalairak C., Panichkitkosolkul W., Silanun K., Theeramunkong T., Nitikraipot S., Suebnukarn S. Forecasting COVID-19 cases using time series modeling and association rule mining. *BMC Med. Res. Methodol.* 2022; 22 (1): 281. https://doi.org/10.1186/s12874-022-01755-x

27. Talkhi N., Akhavan Fatemi N., Ataei Z., Jabbari Nooghabi M. Modeling and forecasting number of confirmed and death caused COVID-19 in IRAN: A comparison of time series forecasting methods. *Biomed Signal. Process Control.* 2021; 66: 102494. https://doi.org/10.1016/j.bspc.2021.102494

28. Tan C.V., Singh S., Lai C.H., Zamri A.S.S.M., Dass S.C., Aris T.B., Ibrahim H.M., Gill B.S. Forecasting COVID-19 Case Trends Using SARIMA Models during the Third Wave of COVID-19 in Malaysia. *International Journal of Environmental Research and Public Health.* 2022; 19 (3): 1504. https://doi.org/10.3390/ijerph19031504

29. Wang M., Wang H., Wang J., Liu H., Lu R., Duan T., Gong X., Feng S., Liu Y., Cui Z., Li C., Ma J. A novel model for malaria prediction based on ensemble algorithms. *PLoS One.* 2019; 14 (12): e0226910. https://doi.org/10.1371/journal.pone.0226910

30. Wang M., Pan J., Li X., Li M., Liu Z., Zhao Q., Luo L., Chen H., Chen S., Jiang F., Zhang L., Wang W., Wang Y. ARIMA and ARIMA-ERNN models for prediction of pertussis incidence in mainland China from 2004 to 2021. *BMC Public Health.* 2022; 22 (1): 1447. https://doi.org/10.1186/s12889-022-13872-9

31. Wang T., Liu J., Zhou Y., Cui F., Huang Z., Wang L., Zhai S. Prevalence of hemorrhagic fever with renal syndrome in Yiyuan County, China, 2005-2014. *BMC Infect Dis*. 2016; 16: 69. https://doi.org/10.1186/s12879-016-1404-7

32. Wang Y.W., Shen Z.Z., Jiang Y. Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China. *PLoS One.* 2018; 13 (9): e0201987. https://doi.org/10.1371/journal.pone.0201987

33. Yan C.Q., Wang R.B., Liu H.C., Jiang Y., Li M.C., Yin S.P., Xiao T.Y., Wan K.L., Rang W.Q. Application of ARIMA model in predicting the incidence of tuberculosis in China from 2018 to 2019. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2019; 40 (6): 633-637. https://doi.org/10.3760/cma.j.issn.0254-6450

34. Zhussupov B., Saliev T., Sarybayeva G., Altynbekov K., Tanabayeva S., Altynbekov S., Tuleshova G., Pavalkis D., Fakhradiyev I. Analysis of COVID-19 pandemics in Kazakhstan. *J. Res. Health Sci.* 2021; 21 (2): e00512. https://doi.org/10.34172/jrhs.2021.52

Received 29.10.2025
Sent for revision 11.11.2025
Accepted 28.11.2025
Published online 30.06.2026

*М. А. Сорокина[1], И. В. Коршуков[1], Н. К. Омарбекова[1], Д. Т. Алибиева[1], Е. А. Дробченко[1]**

КРАТКОСРОЧНОЕ ПРОГНОЗИРОВАНИЕ С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ SARIMA НА ПРИМЕРЕ ЭПИДЕМИОЛОГИЧЕСКИХ ДАННЫХ В ВИДЕ ВРЕМЕННЫХ РЯДОВ ПО COVID-19 В КАЗАХСТАНЕ

[1]Кафедра информатики и биостатистики, НАО «Карагандинский медицинский университет» (100008, Республика Казахстан, г. Караганда, ул. Гоголя, 40; e-mail: info@qmu.kz)

***Елена Александровна Дробченко** – Кафедра информатики и биостатистики, НАО «Карагандинский медицинский университет»; 100008, Республика Казахстан, г. Караганда, ул. Гоголя, 40; e-mail: Drobchenko@qmu.kz

Прогнозирование временных рядов является ключевым инструментом в различных областях, особенно в эпидемиологии, где точные прогнозы могут определять распределение ресурсов и информировать о стратегиях общественного здравоохранения. В данном исследовании изучается применение моделей сезонной авторегрессии с интегрированным скользящим средним (SARIMA) для краткосрочного прогнозирования временных рядов на основе эпидемиологических данных по COVID-19 из Казахстана в качестве практического примера. Набор данных охватывает общее количество подтвержденных случаев, ежедневных пациентов, получающих амбулаторную помощь, и ежедневных пациентов, госпитализированных, за период с 30 мая 2022 г. по 14 декабря 2022 г., с прогнозами на 10 дней вперед, до 24 декабря 2022 г. Анализ использует способность SARIMA фиксировать как сезонные, так и нестационарные закономерности, что делает его высокоэффективным инструментом для моделирования сложной эпидемиологической динамики.

Методология включает несколько ключевых этапов: предварительную обработку данных посредством преобразования в натуральный логарифм для стабилизации дисперсии, проверку стационарности с использованием графиков автокорреляционной функции (АКФ) и дифференцирование для устранения нестационарности. Оптимальными моделями SARIMA были определены (5,3,1)(1,1,0)_7 для всех случаев, (1,2,2)(0,1,1)_7 для пациентов амбулаторного лечения и (2,2,1)(2,1,1)_7 для пациентов госпитализированных с точностью прогнозирования 99%, 97% и 91% соответственно. Эта высокая точность, измеренная с помощью средней абсолютной процентной ошибки (MAPE), подчеркивает надежность SARIMA в краткосрочных прогнозах. Анализ остатков, включая тесты Шапиро-Уилка и Льюнга-Бокса, подтвердил нормальное распределение и независимость остатков моделей, что подтверждает их пригодность для прогнозирования.

Данное исследование подчtркивает универсальность модели SARIMA в отслеживании еженедельных сезонных закономерностей, наблюдаемых в 7-дневных циклах данных о COVID-19, которые отражают тенденции в отчётности или поведении. Хотя пример сосредоточен на эпидемиологических данных, методология широко применима и в других областях, таких как экологический мониторинг или экономическое прогнозирование, где преобладают сезонные временные ряды. Результаты показывают, что модели SARIMA обеспечивают надtжную основу для краткосрочного прогнозирования, предоставляя практические рекомендации для мер вмешательства в сфере общественного здравоохранения и планирования ресурсов, а набор данных о COVID-19 служит наглядным примером эффективности и адаптивности этого подхода.

Ключевые слова: прогнозирование временных рядов; модели SARIMA; краткосрочное прогнозирование; эпидемиологические данные; сезонные закономерности; ARIMA; Казахстан (в качестве примера)

*М. А. Сорокина[1], И. В. Коршуков[1], Н. К. Омарбекова[1], Д. Т. Алибиева[1], Е. А. Дробченко[1]**

SARIMA МОДЕЛЬДЕРІН ПАЙДАЛАНЫП, ҚЫСҚА МЕРЗІМДІ БОЛЖАУ: ҚАЗАҚСТАНДАҒЫ COVID-19 БОЙЫНША ЭПИДЕМИОЛОГИЯЛЫҚ УАҚЫТ ҚАТАРЫНЫҢ ДЕРЕКТЕРІНЕ ҚОСЫМША

[1]«Қарағанды медицина университеті» ҚеАҚ информатика және биостатистика кафедрасы (100008, Қазақстан Республикасы, Қарағанды қ., Гоголь к., 40; e-mail: info@qmu.kz)

***Елена Александровна Дробченко** – «Қарағанды медицина университеті» ҚеАҚ информатика және биостатистика кафедрасы; 100008, Қазақстан Республикасы, Қарағанды қ., Гоголь к., 40; e-mail: Drobchenko@qmu.kz

Уақытша қатарларды болжау әртүрлі салаларда, әсіресе эпидемиологияда негізгі құрал болып табылады, мұнда нақты болжамдар ресурстарды бөлуге және денсаулық сақтау стратегияларын хабарлауға мүмкіндік береді. Бұл зерттеу тәжірибелік мысал ретінде Қазақстандағы COVID-19 эпидемиологиялық деректерін пайдалана отырып, қысқа мерзімді уақыт қатарын болжау үшін маусымдық авторегрессивті біріктірілген жылжымалы орташа (SARIMA) үлгілерін қолдануды зерттейді. Деректер жинағы 2022 жылдың 30 мамырынан бастап 14 желтоқсанына дейінгі аралықтағы расталған жағдайлардың, күнделікті амбулаториялық науқастардың және күнделікті ауруханаға жатқызылған науқастардың жалпы санын қамтиды, болжамдармен 10 күн бұрынғы, 2022 жылдың 24 желтоқсанына дейін. Талдау жасау SARIMA-ның маусымдық және стационарлық емес үлгілерді түсіру мүмкіндігін пайдаланады, бұл оны күрделі эпидемиологиялық динамикаларды модельдеу үшін жоғары тиімді құрал етеді.

Әдістеме бірнеше негізгі қадамдарды қамтиды: дисперсияны тұрақтандыру үшін натурал логарифмге түрлендіру арқылы деректерді алдын ала өңдеу, автокорреляция функциясының (АКФ) графиктерін пайдаланып стационарлықты тексеру және стационарлық еместікті жою үшін дифференциалдау. Анықталған ең тиімді SARIMA үлгілері: барлық жағдайлар үшін (5,3,1)(1,1,0)_7, амбулаториялық науқастар үшін (1,2,2)(0,1,1)_7 және ауруханаға жатқызылған науқастар үшін (2,2,1)(2,1,1)_7 болды. Болжаудың дәлдігі сәйкесінше 99%, 97% және 91% болды. Орташа абсолютті пайыздық қателікпен (MAPE) өлшенетін бұл жоғары дәлдік SARIMA-ның қысқа мерзімді болжамдардағы сенімділігін көрсетеді. Қалдықтарды талдау, соның ішінде Шапиро-Вилк және Льюнг-Бокс сынақтары, үлгі қалдықтарының қалыпты тарамдалуын және тәуелсіздігін, сонымен қатар олардың болжам жасау үшін жарамдылығын растады.

Бұл зерттеу есеп беру немесе мінез-құлық үрдістерін көрсететін COVID-19 деректерінің 7 күндік циклдарында байқалатын апталық маусымдық үлгілерді түсірудегі SARIMA моделінің әмбебаптығын көрсетеді. Демек, мысал эпидемиологиялық деректерге бағытталғанымен, әдістеме маусымдық уақыт қатарлары басым болатын экологиялық мониторинг немесе экономикалық болжау сияқты басқа салаларға кеңінен қолданылады. Алынған нәтижелер SARIMA үлгілері қоғамдық денсаулық сақтау шаралары мен ресурстарды жоспарлау үшін тәжірибелік нұсқаулар беретін қысқа мерзімді болжау үшін сенімді негізді қамтамасыз ететінін және COVID-19 деректер жинағы осы тәсілдің тиімділігі мен бейімделуінің айқын мысалы ретінде қызмет ететінін көрсетеді.

Кілт сөздер: уақыттық қатарларды болжау; SARIMA үлгілері; қысқа мерзімді болжау; эпидемиологиялық деректер; маусымдық үлгілер; ARIMA; Қазақстан (мысал ретінде)