UDC 616-036.22.001.57:616.9-022:578.834.1
DOI: 10.20535/SRIT.2308-8893.2024.4.02

SHORT-TERM FORECASTING OF THE MAIN INDICATORS OF THE COVID-19 EPIDEMIC IN UKRAINE BASED ON THE SEASONAL CYCLE MODEL

A.B. ALYOKHIN, A.B. BRUTMAN, A.N. GRABOVOY, T.V. SHABELNYK

Abstract. The authors of this study propose a method of short-term forecasting of time series of the main indicators of the COVID-19 epidemic, which has a pronounced seasonality. This method, which has no direct analogies, provides the decomposition of a general forecasting task into several simpler tasks, such as the tasks of building a model of the seasonal cycle of a time series, aggregating the original time series, taking into account the duration of the seasonal cycle, forecasting an aggregated time series, developing an aggregated forecast into a forecast in the original time scale, using the seasonal cycle model. The solution for each task allows the usage of relatively simple methods of mathematical statistics. The article provides a formally rigorous description of all procedures of the method and illustrations of their numerical implementation on the example of a real forecasting task. The use of this method for short-term forecasting of the COVID-19 epidemic development in Ukraine has systematically demonstrated its effectiveness.

Keywords: COVID-19 epidemic, time series, short-term forecasting, seasonal cycle, indicators.

INTRODUCTION

Despite the fact that the first works on mathematical epidemiology appeared as early as the 20s of the last century [1], the COVID-19 pandemic, especially in the initial phase, created significant difficulties in building high-quality forecasts of the disease spread based on mathematical models [2–4]. This, along with the enormous scale of the pandemic, as well as its socio-economic consequences, has attracted the attention of scientists around the world to the problem of quantitative forecasting of the development of the COVID-19 epidemic.

Currently, a wide variety of mathematical tools are used to model and forecast the spread of COVID-19, among which the main place is occupied by systemic models of epidemics, including simulation, statistical models and methods for forecasting time series, methods and models of artificial intelligence (neural networks and etc.).

LITERATURE ANALYSIS AND PROBLEM STATEMENT

Taking into account the pronounced seasonality of COVID-19 epidemic development in various countries of the world, the Box–Jenkins (SARIMA) [5] and Holt-Winters [6] methods, which take into account seasonal effects, are the most widely used for statistical modelling of epidemics. SARIMA models are a linear combination of series and seasonal profile elements, as well as past forecast errors considering autoregression. The Holt-Winters method uses the technique of triple exponential smoothing applied to the main components of a statistical time series: series level, trend, seasonal component. Both models have their own advantages and disadvantages.

This, in the absence of the only best approach or forecasting model, taking into account the principle of multiple models generally recognized in forecasting theory, leaves room for development and determines the relevance of efforts aimed at developing alternative approaches to forecasting which more accurately reflect the specifics of the forecasting object, including features of seasonal cyclicity, increasing the level of transparency, formalization and simplicity of the procedures applied.

The paper proposes a new method that decomposes the general problem of forecasting into a number of simple problems. These are the tasks of building a model of the seasonal cycle of a time series, aggregating the original time series taking into account the duration of the seasonal cycle, forecasting the aggregated time series and transforming the aggregated forecast into a forecast in the original time scale using the seasonal cycle model.

PURPOSE OF THE STUDY

The aim of the study is to develop a "direct" method for short-term forecasting of time series of the main indicators of the COVID-19 epidemic based on a seasonal cycle model.

MATERIALS AND METHODS OF THE STUDY

The study is based on statistical data from the Public Health Center of the Ministry of Health of Ukraine [7], as well as indicators derived from them, which together characterize the spread of the COVID-19 epidemic in Ukraine over the entire observation period. This statistical data is a multidimensional discrete time series with a 1-day increment, which includes univariate statistical series of three basic indicators (increase in infected, deaths, recoveries), five derived daily and cumulative indicators (increase in active cases; total number of infected, deaths, recoveries, number of active cases) and three synthetic derived indicators (disease spread and mortality, epidemic progression).

To determine the availability and duration of a cycle in the statistical series of reference indicators, autocorrelation methods were used, implemented in the statistical software IBM SPSS Statistics, STATISTICA and MS Excel 2019. For aggregating time series, the corresponding series conversion procedures of the specified application program packages were used.

The building of cycle models, as well as trend models of aggregated time series of reference indicators, was carried out using generally accepted methods of analytical alignment of time series and, in particular, curve-fitting methods using non-linear optimization methods.

To develop and evaluate indicative forecasts of reference indicators, statistical series were used that included data from the 36th to the 57th week of the entire observation period. Forecasts were built for a period of $1+1$ weeks, where the forecast for the second week is an estimate. Statistical series of reference indicators from the 36th to the 55th week were used as training sequences, and the next two weeks were used as a verification (forecasting) sequence. The accuracy of forecasts for all forecasted indicators was assessed using the calculated (forecasting) data and data from verification sequences using the MAPE accuracy indicator.

MATHEMATICAL MODELS

In the general scheme of the developed method for short-term forecasting of the COVID-19 epidemic progress, for each one-dimensional statistical series of the reference indicator, the following procedures are performed:

– identification of the seasonal cycle in the time series and determination of its duration;

– building of a general model of the seasonal cycle (seasonal profile of the time series) or a set of models of seasonal cycles of the time series;

– removal of the seasonal component by aggregating the original time series with a step equal to the cycle duration, and building a trend model of the aggregated series;

– trend forecasting for the forecast period in the aggregated time scale;

– disaggregation of the aggregated forecast using the cycle model (cycle models), which is the decomposition of the aggregated forecast into a forecast in the original time scale.

The calculation of the forecast values of all derived indicators of the COVID-19 epidemic is based on the forecasts of the reference indicators.

For a formal description of the above stated procedures, we introduce the following notation:

$\Re_S$ — original multidimensional statistical time series;

n — dimension of the time series $\Re_S$;

I — set of indices of the components of the original time series, $I=\{1,\dots,n\}$;

t — discrete moment of time;

T_s — length of the time series $_S$;

T_F — duration of the forecasting period;

$\Re_F$ — forecasting time series of length T_F;

Let R — n-dimensional vector of COVID-19 epidemic indicators. Then

$$R=(R^*,R',R''),$$

where R^* — vector of epidemic reference indicators of dimension n^*; R' — vector of daily and cumulative derivative indicators of the epidemic of dimension n'; R'' — vector of synthetic derivative indicators of dimension n''.

In this study, the following assumptions have been made:

$$R^*\{\Delta TC,\Delta D,\Delta R\},$$

where $\Delta TC, \Delta D, \Delta R$ — daily increase of infected, deaths, recovered respectively.

$$R' \in \{\Delta AC, TC, D, R, AC\},$$

where ΔAC — daily increase in the number of active cases at time *t*; TC, D, R — total number of infected, deaths and recovered; AC — number of active cases.

$$R'' \in \{R_t, I_{TC}, I_{CC}, I_P\},$$

where R_t — infection spread rate; I_{TC} — fatality rate (according to the number of infected (total cases)); I_{CC} — fatality rate (according to the number of closed cases); I_P — epidemic progress indicator.

The following notations refer to an arbitrary one-dimensional time series $\Re_S(i)$, where $i \in I^*$, where I^* — a subset of epidemic reference indices $(|I^*| = n^*)$. For simplicity, the index *i*, where it does not generate ambiguity, will be omitted below:

T_C — length of the seasonal cycle time series $\Re_S(i)$;

k — number of complete cycles in time series $\Re_S(i)$;

t' — sequence number of the seasonal cycle or, what is the same, a discrete point in time in the aggregated time scale;

t'' — sequence number of the observation in the cycle;

$\Re_{SC}(t')$ — t'-th statistical time series of length T_C (series of seasonal cycle elements t').

$\Re'_S$ — aggregated statistical time series of observations;

T'_S — length of the aggregated time series $\Re'_S$ in the aggregated time scale (in units of the number of complete cycles, $T'_S = k$);

T'_F — duration of the forecast period in the aggregated time scale.

$\Re_{SCN}(t')$ — t'-th normalized seasonal cycle (normalized cycle time series);

M_{CN} — general model of normalized seasonal cycle (seasonal profile of time series $\Re_S(i)$ — model time series of length T_C).

Considering the introduced notation, the formal description of all stages of forecasting of the reference one-dimensional series $\Re_S(i)$, $i \in I^*$, is as follows.

Stage 1. Analysis of the time series cyclicity.

1.1. Determining the availability, duration of T_C and the number *k* of complete cycles in the original time series $\Re_S(i)$.

This stage is implemented by standard methods of autocorrelation analysis.

Stage 2. Building of the normalized cycle model $M_{CN}(i)$ of the time series.

2.1. Normalization of the levels of the seasonal cycle time series $\Re_{SC}(i, t')$ for all t', $t'\{1,\ldots,T'_S\}$.

Normalization is carried out according to the formula:

$$r_{CN}(i,t',t") = \frac{r_{CS}(i,t',t")}{\sum_{t"} r_{CS}(i,t',t")} \quad \forall t',$$

where $r_{CS}(i,t',t")$ — $t"$-th element of the t'-th cycle of the time series $\Re_S(i)$; $r_{CN}(i,t',t")$ — $t"$-th element of the t'-th cycle of the time series $\Re'_S(i)$.

2.2. Building of a mathematical model of the normalized cycle $M_{CN}(i)$ based on T_S observations — the set of all cycles $\{\Re'_{SCN}(i,t')\}$.

The seasonal cycle model $M_{CN}(i)$ is developed on the elements of the normalized time series of cycles $\{\Re_{SCN}(t')\}$, the set of which acts as a repetitiveness. In the study, a degree l polynomial from the sequence number of the cycle element was used as a model of the seasonal cycle, where $l < T_C$.

$$M_{CN}(i.t'') = a_l\,(t'')^l + a_{l-1}(t'')^{l-1} + \ldots + a_1\,t'' + a_0,$$

where $t''\{1,\ldots,T_c\}$; A — parameter vector of the degree 1 polynomial, $A = (a_0,\ldots,a_l)$.

The parameters A of the $M_{CN}(i)$ model are defined as a solution to an optimization problem of the type:

$$\min_A \sum_{t',t'} \alpha_{t'}(r_{CM}(i,t") - r_{CN}(i,t',t"))^2 \quad \forall t', t' \in \{1,\ldots,k\};$$

$$r_{CM}(i,t") = a_l\,(t'')^l + a_{l-1}\,(t'')^{l-1} + \ldots + a_1\,t'' + a_0 \quad \forall t"; \quad 0 \le r_{CM}(i,t") \le 1 \quad \forall t",$$

$r_{CM}(i,t")$ — $t"$-th model cycle element $M_{CN}(i)$; α_t — significance coefficient of the t'-th cycle.

In the model, the weight coefficients $\{\propto_{t'}\}_{t'}$ are set using the logistic function of the cycle t' sequence number

Stage 3. Building a model of the aggregated series $\Re'_S(i)$.

3.1. Aggregation of the original one-dimensional time series $\Re_S(i)$ — formation of an aggregated time series $\Re'_S(i)$ of length T'_S.

The aggregation of the series $\Re_S(i)$ is carried out with a step equal to the duration of the cycle T_C, using the operation of summing as the aggregation operation.

3.2. Building a trend model $M_T(i)$ of the aggregated time series $\Re'_S(i)$.

It is carried out by an arbitrary method of analytical alignment of time series.

Stage 4. Forecasting of the initial time series $\Re_S(i)$.

4.1. Development of an aggregated forecast $\Re'_F(i)$ for the forecast time T_F.

It is carried out by extrapolating the trend using the $M_T(i)$ trend model for the forecast period T_F in the aggregated time scale.

4.2. Formation of the forecasting time series $\Re_F(i)$ based on the aggregated forecast $\Re'_F(i)$.

At this stage, deconvolution of each element $r'_F(i,t')$ of the forecasting series $\Re'_F(i)$ is carried out using the $M_{CN}(i)$ cycle model, as a result, each element is replaced by the time series of the corresponding cycle *t'* in the original time series scale, the elements of which are determined by the following formula:

$$r_F(i,t) = r'_F\left(i,\left[\frac{t}{T_C}\right]+1\right)\times \times M_{CN}\left(i,\left(t-\left[\frac{t}{T_C}\right]\times T_C\right)\right) \quad \forall t,\ t\in\{1,\ldots,T'_F\times k\};$$

where $[a]$ — an integer part of the number *a*.

As a result of this stage, the aggregated forecast series $\Re'_F(i)$ in the aggregated time scale is disaggregated into the forecast series $\Re_F(i)$ in the original time scale.

After forecasting of all COVID-19 epidemic reference indicators, the forecast values of the epidemic derived indicators are calculated in accordance with the formulas below.

Stage 5. Calculation of the main derivatives (synthetic) indicators of the COVID-19 epidemic

The daily increase in active cases ΔAC at each time *t* is derived from the increase in infected ΔTC, deaths ΔD and recovered ΔR and is calculated by the following formula:

$$\Delta AC_t = \Delta TC_t - \Delta D_t - \Delta R_t \quad \forall t$$

Cumulative indicators TC, D, R, AC are calculated using the following formulas:

$$P_t = P_{t-1} + \Delta P_t \quad \forall P,\ P\in\{TC,\ D,\ R,\ AC),\ t\,.$$

The I_{TC} fatality rate is calculated using the formula:

$$I_{TC} = D_t / TC_t \quad \forall t\,.$$

and the I_{CC} fatality rate — according to the formula:

$$I_{CC} = D_t / (D_t - R_t) \quad \forall t$$

As one of the meaningful tools for monitoring and analysing the development of the COVID-19 epidemic, the authors propose an I_P progress indicator, which is the ratio of fatality rates:

$$I_P(t) = I_{CC}(t) / I_{TC}(t) = (D_t + R_t) / TC_t \quad \forall t.$$

A physical meaning of this indicator is transparent, reflected in its name and indicates the traversed path in % of the epidemic progress from the moment of its emergence to the moment of completion.

Not least informative is the statistical analogue of the reproduction coefficient R_0 — the infection spread coefficient R_t, which is proposed to be calculated taking into account the seasonality of the time series:

$$R_t = \sum_{l=t-T_c+1}^{t} R_l \Big/ \sum_{m=t-2T_c+1}^{t-T_c} R_m \quad \forall t\,.$$

RESULTS AND DISCUSSION

Let us illustrate the implementation of each of the above mentioned stages, using the example of short-term forecasting of the development of the COVID-19 epidemic in Ukraine.

Stage 1. Analysis of the time series cyclicity of the reference indicator.

In the study, to assess the availability and duration of seasonality in these series, the FORECAST.ETS.SEASONALITY function contained in MS Excel 2019 was used, which made it possible to detect a seasonal cycle equal to 7 days (see Fig. 1).

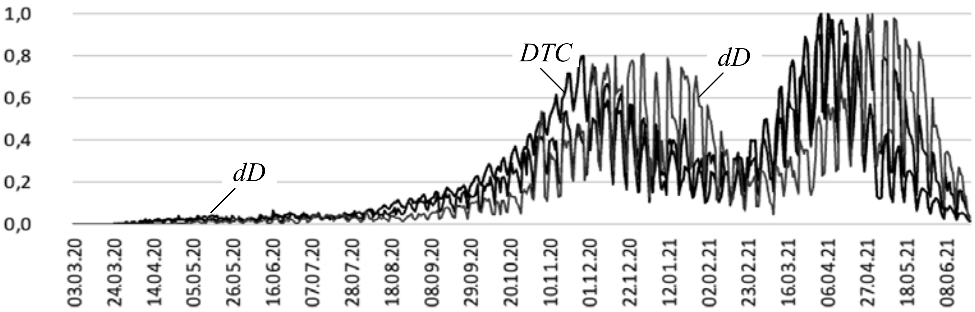


Fig. 1. Daily increase in Infected (ΔTC), Deaths (ΔD) and Recovered (ΔR) from COVID-19 in Ukraine (normalized data)

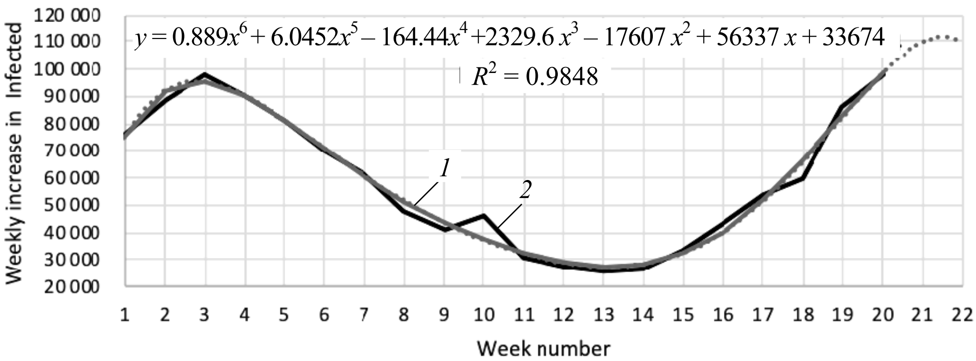
The proposed method will be further illustrated for the time series of the indicator of the daily increase in infected (ΔTC).

Stage 2. Building a seasonal cycle model.

As a model of the weekly seasonal profile, a degree 4 polynomial was used, the parameters of which are determined by solving the corresponding nonlinear optimization problem. In doing so, the weight coefficients of the actual weekly cycles were set using the logistic function of the serial number of the cycle.

Stage 3. Building a trend model of an aggregated time series.

In this study, a polynomial no higher than a degree 6 is used as a model of the aggregated series for illustrative purpose (Fig. 2).



Stage 4. Forecasting the initial time series.

Extrapolation of the aggregated time series is carried out by substituting the number of the forecast period into the equation of the $y = f(x)$ (see Fig. 2), and disaggregation of the forecast — by deconvolution of the obtained values using the cycle model.

The results of these procedures for all reference indicators ΔTC, ΔD and ΔR, as well as the daily increase in active cases ΔAC are shown in Fig. 3.

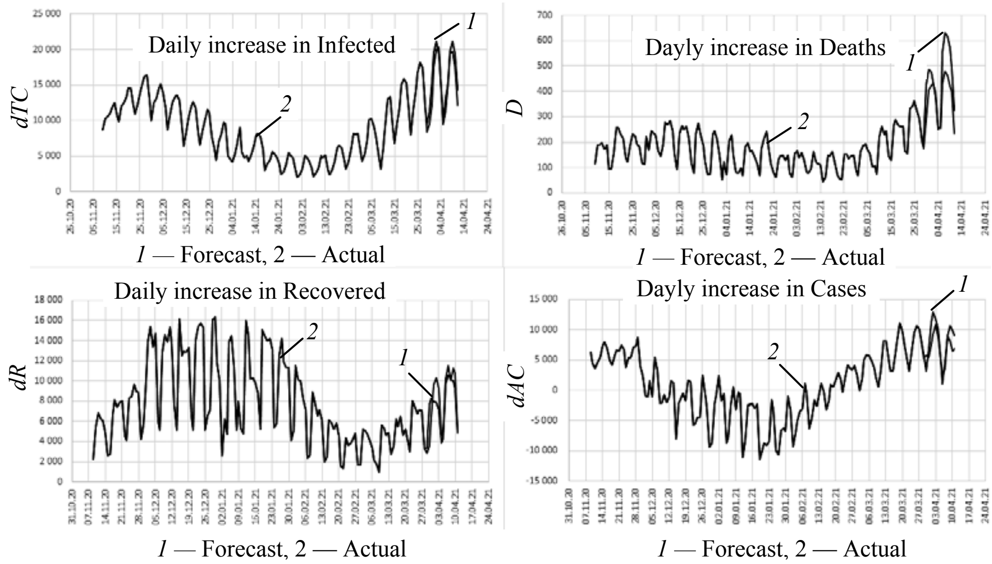


Fig. 3. Actual and forecast values of daily indicators of the COVID-19 epidemic in Ukraine

Step 5. Calculation of derived indicators of the COVID-19 epidemic.

The forecast values of derived indicators are calculated in accordance with the formulas described above. Their correspondence to the actual data is shown in Fig. 4.

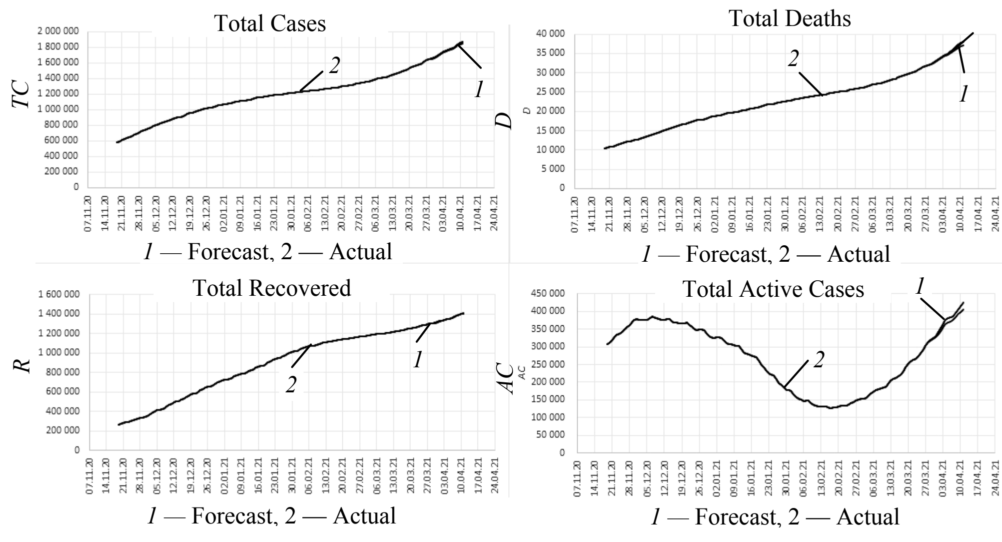


Fig. 4. Actual and forecast values of cumulative indicators of the COVID-19 epidemic in Ukraine

Evaluation of forecast accuracy. In order to evaluate the performance of the proposed method, the accuracy of forecasts has been assessed for the main indicators of the COVID-19 epidemic in Ukraine using the mean absolute percentage error MAPE, as well as a comparative analysis of the accuracy of this method with the exponential smoothing methods and SARIMA mentioned at the beginning of the paper (see Table).

As follows from the data in Table, the estimation results confirm the fact, well known in statistical forecasting of time series, that different methods deal with different situations in different ways, and this fact does not allow us to give preference to one of them in general.

Estimates (MAPE) of the forecast accuracy of the main indicators COVID-19 epidemic in Ukraine, %

N	Indicator	Weekly forecast			2-week forecast		
		M_1	M_2	M_3	M_1	M_2	M_3
1	TC	0.56	0.17	0.14	0.60	0.30	0.10
2	D	0.66	1.01	0.74	1.27	1.17	0.92
3	R	0.21	0.48	0.31	0.22	1.13	0.65
4	AC	1.91	2.00	2.00	2.76	3.22	2.88
5	ΔTC	12.30	19.28	11.19	10.48	12.83	8.59
6	ΔD	14.51	10.66	13.07	21.03	8.44	10.02
7	ΔR	34.28	35.27	39.06	21.20	36.06	29.90
8	ΔAC	27.70	16.41	17.19	28.85	34.97	31.14
9	I_{TC}	0.16	1.02	0.88	0.70	0.96	1.00
10	I_{CC}	0.45	0.52	0.42	1.26	0.55	0.35
11	I_{P}	0.34	0.50	0.46	0.57	0.92	0.74

where M_1 — MAPE of proposed method; M_2 — MAPE of exponential smoothing; M_3 — MAPE of SARIMA.

As follows from the data in Table 1, the estimation results confirm the fact, well known in statistical forecasting of time series, that different methods deal with different situations in different ways, and this fact does not allow us to give preference to one of them in general.

It should also be noted that the procedures that implement these methods in the statistical software IBM SPSS Statistics provide for the generation of a series of models and the selection of the best among them. In assessing the accuracy of the method proposed in this study base forecast models were used, and the potential for improvement of forecast models, which is inherent in the method, is not utilized. The task of realizing this potential to improve the efficiency of the method is the subject of further research.

The corresponding calculations were carried out on a systematic basis within the framework of the activities of the "Working Group on Mathematical Modelling of Problems Related to the SARS-CoV-2 Coronavirus Epidemic in Ukraine" of the National Academy of Sciences of Ukraine during 2020 and confirm the method effectiveness proposed by the authors of the study.

CONCLUSIONS

The time series of daily growth indicators of the COVID-19 epidemic have a pronounced seasonal pattern.

The method proposed by the authors makes full use of this circumstance and implements the idea of decomposing the general task of developing short-term forecasts of the main indicators of the COVID-19 epidemic into a number of particular subtasks, for which simple methods of mathematical statistics are applicable. In particular, the “direct” method of identifying the seasonal cycle makes it possible to use quite simple mathematical models of an arbitrary form to describe the seasonal profile of a time series. Aggregation of the original time series with a step equal to the seasonal cycle duration enables to eliminate the seasonal component without distorting the information, and the problem of forecasting the initial time series can be transformed into the simpler problem of extrapolating the trend model of the aggregated time series with subsequent deconvolution (using the cycle model) of its forecasting values in daily values of the corresponding indicators.

Practical use of this method for developing short-term forecasts of the COVID-19 epidemic progress in Ukraine on a systematic basis has demonstrated quite satisfactory accuracy of forecasts (MAPE estimates).

REFERENCES

1. W.O. Kermack, A.G. McKendrick, “A Contribution to the Mathematical Theory of Epidemics,” *Proc. Roy. Soc. Lond. A*., no. 115, pp. 700–721, 1972.
2. S. Borenstein, C.K. Johnson, “Modeling coronavirus: ‘Uncertainty is the only certainty,’” *AP News*, April 8, 2020. Accessed on: February 10, 2022. [Online]. Available: https://apnews.com/article/public-health-health-us-news-ap-top-news-virus-outbreak-88866498ff5c908e5f28f7b5b5e5b695
3. V. Romanenko, Y. Milyavsky, “Combined control of multirate impulse processes in a cognitive map of COVID-19 morbidity,” *System Research and Information Technologies*, no. 3, pp. 46–56, 2022. doi: https://doi.org/10.20535/SRIT.2308-8893.2022.3.03
4. L Guryanova, R. Yatsenko, N. Dubrovina, V. Babenko, and V. Gvozditskyi, “Machine learning methods and models, predictive analytics and applications: development trends in the post-crisis syndrome caused by COVID-19,” *CEUR Workshop Proceedings*, 2927, pp. 1–7, 2021.
5. G.E.P. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung, *Time Series Analysis: Forecasting and Control*. New Jersey: John Wiley & Sons, Inc., 2015, 712 p.
6. P.R. Winters, “Forecasting sales by exponentially weighted moving averages” *Management Science*, no. 6, pp. 324–342, 1960.
7. “Koronavirusna infektsiia COVID-19 [Corona virus infection COVID-19],” *Tsentr hromadskoho zdorovia MOZ Ukrainy*. Available: https://phc.org.ua/kontrol-zakhvoryuvan/inshi-infekciyni-zakhvoryuvannya/koronavirusna-infekciya-covid-19
8. “Prohnoz rozvytku epidemii COVID-19 v Ukraini na period 7–13 hrudnia 2020 roku vid uchenykh Instytutu problem rynku ta ekonomiko-ekolohichnykh doslidzhen NAN Ukrainy [Forecast of the development of the COVID-19 epidemic in Ukraine for the period of December 7–13, 2020 by scientists of the Institute of Market Problems and Economic and Environmental Research of the National Academy of Sciences of Ukraine],” *Natsionalna akademiia nauk Ukrainy*, December 7, 2020. Available: https://www.nas.gov.ua/ UA/Messages/Pages/View.aspx?MessageID=7237

Received 30.08.2023

INFORMATION ON THE ARTICLE

Alexei B. Alyokhin, ORCID: 0000-0001-5209-8036, Mariupol State University, Ukraine, e-mail: aba99@ukr.net

Anna B. Brutman, ORCID: 0000-0002-7774-5356, National University "Zaporizhzhia Polytechnic", Ukraine, e-mail: a_brutman@yahoo.com

Alexandr N. Grabovoy, ORCID: 0000-0001-5705-9909, Bogomolets National Medical University, Ukraine, e-mail: angrabovoy@gmail.com

Tetiana V. Shabelnyk, ORCID: 0000-0001-9798-391X, Simon Kuznets Kharkiv National University of Economics, Ukraine, e-mail: tanya.shabelnik17@gmail.com

КОРОТКОСТРОКОВЕ ПРОГНОЗУВАННЯ ОСНОВНИХ ПОКАЗНИКІВ ЕПІДЕМІЇ В УКРАЇНІ НА ОСНОВІ МОДЕЛІ СЕЗОННИХ ЦИКЛІВ / О.Б. Альохін, А.Б. Брутман, О.М. Грабовий, Т.В. Шабельник

Анотація. Запропоновано метод короткострокового прогнозування часових рядів основних показників епідемії COVID-19, яким притаманна виражена сезонність. Зазначений метод, що не має прямих аналогів, передбачає декомпозицію загального завдання прогнозування на ряд більш простих завдань, таких як побудова моделі сезонного циклу часового ряду, агрегування вихідного часового ряду з урахуванням тривалості сезонного циклу, прогнозування агрегованого часового ряду, розгортання агрегованого прогнозу в прогноз у вихідній часовій шкалі за допомогою моделі сезонного циклу, вирішення кожного з яких допускає застосування відносно простих методів математичної статистики. Наведено формально строге описання всіх процедур методу та ілюстрації їх числової реалізації на прикладі реального завдання прогнозування. Застосування зазначеного методу для розроблення короткострокових прогнозів розвитку епідемії COVID-19 в Україні на систематичній основі продемонструвало його ефективність.

Ключові слова: епідемія COVID-19, часові ряди, короткострокове прогнозування.