

Голубенко Олександр Іванович

Заклад вищої освіти «Міжнародний науково-технічний університет імені академіка Юрія Бугая», Київ, Україна

ORCID ID: 0000-0002-1776-5160

Маковейчук Олександр Миколайович

Заклад вищої освіти «Міжнародний науково-технічний університет імені академіка Юрія Бугая», Київ, Україна

ORCID ID: 0000-0003-4425-016X

Савченко Андрій Васильович

Заклад вищої освіти «Міжнародний науково-технічний університет імені академіка Юрія Бугая», Київ, Україна

ORCID ID: 0000-0002-8314-6034

Власенко Вадим Олександрович

Державний університет інформаційно-комунікаційних технологій, Київ, Україна

ORCID: 0000-0002-9329-5914

v.vlasenko@duikt.edu.ua

ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ З ВИКОРИСТАННЯМ ARIMA-МОДЕЛЕЙ

Анотація: У статті досліджено підходи до прогнозування чисельності населення України із використанням ARIMA-моделей. Проаналізовано класичний підхід Бокса-Дженкінса для побудови моделей ARIMA різного порядку. Наведено результати моделювання за даними Державної служби статистики України за період 1990–2019 років із подальшим прогнозуванням чисельності населення на 2020 рік. Порівняно точність різних моделей за середньоквадратичним відхиленням (СКВ) та абсолютною різницею прогнозу з фактичними даними. Запропоновано критерії мінімізації СКВ для вибору оптимальної моделі прогнозування.

У дослідженні оцінюються дванадцять моделей ARIMA з різними комбінаціями параметрів авторегресії (p), диференціації (d) та ковзного середнього (q). Вибір оптимальної моделі базувався на двох основних критеріях: середньоквадратичній похибці (RMSE) точності прогнозу та абсолютній різниці між прогнозованими та фактичними значеннями на 2020 рік. У дослідженні висвітлено переваги та обмеження моделей ARIMA у короткостроковому прогнозуванні чисельності населення та запропоновано використання мінімізації RMSE як надійного критерію вибору моделі. Результати дослідження підтверджують ефективність моделей ARIMA у завданнях демографічного прогнозування, які є важливими для прийняття рішень на основі даних у державному управлінні, міському розвитку та соціальній політиці.

Ключові слова: часові ряди, прогноз чисельності населення, модель ARIMA, метод Бокса-Дженкінса, критерії вибору моделі.

Golubenko Oleksandr

Higher Education Institution «Academician Yuriy Bugay International Scientific and Technical University», Kyiv, Ukraine

ORCID ID: 0000-0002-1776-5160

Makoveichuk Oleksandr

Higher Education Institution «Academician Yuriy Bugay International Scientific and Technical University», Kyiv, Ukraine

ORCID ID: 0000-0003-4425-016X

Savchenko Andrii

Higher Education Institution «Academician Yuriy Bugay International Scientific and Technical University», Kyiv, Ukraine

ORCID ID: 0000-0002-8314-6034

Vadym Vlasenko

State University of Information and Communication Technologies, Kyiv, Ukraine

ORCID: 0000-0002-9329-5914

v.vlasenko@duikt.edu.ua

TIME SERIES FORECASTING USING ARIMA MODELS

Abstract: This article explores the application of ARIMA (AutoRegressive Integrated Moving Average) models for forecasting the population of Ukraine. Time series forecasting plays a crucial role in demographic analysis, long-term socio-economic planning, and the development of effective public policies. The study applies the classical Box-Jenkins methodology to build ARIMA models of different orders and assess their forecasting accuracy. The modeling was performed using official data on the population of Ukraine for the period from 1990 to 2019, provided by the State Statistics Service of Ukraine. A population forecast for the year 2020 was generated based on various ARIMA configurations, and the results were compared to actual statistical data.

The study evaluates twelve ARIMA models with different combinations of autoregressive (p), differencing (d), and moving average (q) parameters. The selection of the optimal model was based on two main criteria: the root mean square error (RMSE) of forecast accuracy and the absolute difference between forecasted and actual values for 2020. The research highlights the advantages and limitations of ARIMA models in short-term population forecasting and proposes the use of RMSE minimization as a reliable criterion for model selection. The findings confirm the effectiveness of ARIMA models in demographic forecasting tasks, which are essential for data-driven decision-making in public administration, urban development, and social policy.

Keywords: time series, population forecasting, ARIMA model, Box-Jenkins method, model selection criteria.

Постановка проблеми.

Прогнозування часових рядів є однією з ключових задач сучасного аналізу даних, що має широке застосування у сферах економіки, енергетики, транспорту, фінансів, екологічного моніторингу та інших галузях, де дані накопичуються у вигляді послідовностей спостережень у часі [1], [2]. Надійне прогнозування дозволяє своєчасно реагувати на зміни тенденцій, підвищувати ефективність управлінських рішень, оптимізувати ресурси та знижувати ризики [3].

Особливої актуальності ця проблема набуває у контексті зростання обсягів даних та складності систем, де дані є неповними, містять тренди, сезонність або стохастичні флуктуації [4].

Аналіз останніх досліджень і публікацій.

Серед класичних статистичних методів прогнозування часових рядів особливе місце займають моделі ARIMA (AutoRegressive Integrated Moving Average), що поєднують авторегресію, інтегрування та ковзне середнє для побудови адаптивних прогнозів [5].

Моделі ARIMA відзначаються універсальністю та ефективністю для широкого класу стаціонарних і слабо нестационарних процесів. Завдяки простоті математичної інтерпретації та наявності алгоритмів ідентифікації структури моделі, ARIMA широко застосовуються у прогнозуванні фінансових ринків [6], споживання енергоресурсів [7], динаміки економічних показників [8], а також у системах моніторингу геопросторових даних [9].

Сучасні дослідження [10], [11] підкреслюють перспективність комбінованих підходів, де моделі ARIMA інтегруються з методами машинного навчання для підвищення точності прогнозів. Проте навіть у класичному вигляді ARIMA залишається базовим етапом аналізу часових рядів, зокрема для задач попереднього прогнозування та перевірки гіпотез про природу даних [12].

Мета і задачі дослідження.

Мета роботи полягає в оцінці ефективності застосування ARIMA-моделей різного порядку для прогнозування часових рядів та в обґрунтуванні критеріїв для вибору оптимальної моделі.

Виклад основного матеріалу дослідження.

Модель часової динаміки даних базується на узагальненій моделі стаціонарного випадкового процесу [13]. Випадковий процес називається стаціонарним, якщо в кожен момент часу t випадковий розподіл значень випадкової величини X_t залежить тільки від її попередніх значень і відстані по часу до них, але не від самих значень часу:

$$X_t = c + \sum_{i=1}^p a_i X_{t-i} + \varepsilon_t, \quad (1)$$

де c – деяка константа, можна прийняти не втрачаючи загальності, що $c = 0$; a_i – коефіцієнти моделі; p – порядок моделі; ε_t – білий шум з дисперсією σ^2 , що згенеровано в той самий момент часу t . В такому випадку кажуть, що задана модель AR(p) – авторегресія порядку p [14], [15].

Модель MA(q) ковзного середнього порядку q задається рівнянням [14], [15]:

$$X_t = \mu + \sum_{i=1}^q b_i \varepsilon_{t-i} + \varepsilon_t, \quad (2)$$

де μ є математичним сподіванням X_t (також можна прийняти, що $\mu = 0$).

Узагальнена модель авторегресії-ковзного середнього ARMA(p, q) є об'єднанням моделей авторегресії AR(p) та ковзного середнього MA(q) і задається рівнянням [14]-[16]:

$$X_t = c + \sum_{i=1}^p a_i X_{t-i} + \sum_{i=1}^q b_i \varepsilon_{t-i} + \varepsilon_t, \quad (3)$$

де збережено всі попередні позначення.

Якщо процес X_t не є стаціонарним, то доцільно розглянути $\Delta^d X_t$ – різниці порядку d . Часовий ряд називається інтегрованим порядку d , якщо різниці ряду порядку d є стаціонарними [14]. В такому випадку доцільно використовувати інтегровану модель авторегресії-ковзного середнього ARIMA(p, d, q), яка є узагальненням моделі ARMA(p, q) для різниць ряду $\Delta^d X_t$, тобто:

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{i=1}^q b_i \varepsilon_{t-i} + \varepsilon_t. \quad (4)$$

Одним із недоліків моделі ARIMA(p, d, q) є складність визначення порядку і параметрів моделі. Підхід Бокса-Дженкінса [5] до аналізу часових рядів полягає в тому, що в першу чергу оцінюється стаціонарність ряду. Різними тестами виявляються наявність поодиноких коренів і порядок інтегрованості тимчасового ряду (зазвичай обмежуються першим або другим порядком). Далі при необхідності (якщо порядок інтегрованості більше нуля) ряд перетворюється взяттям різниці відповідного порядку і вже для перетвореної моделі будується деяка ARMA-модель, оскільки передбачається, що отриманий процес є стаціонарним, на відміну від вихідного нестаціонарного процесу (різницево-стаціонарного або інтегрованого процесу порядку d). Схематично цей процес представлено на рис. 1.

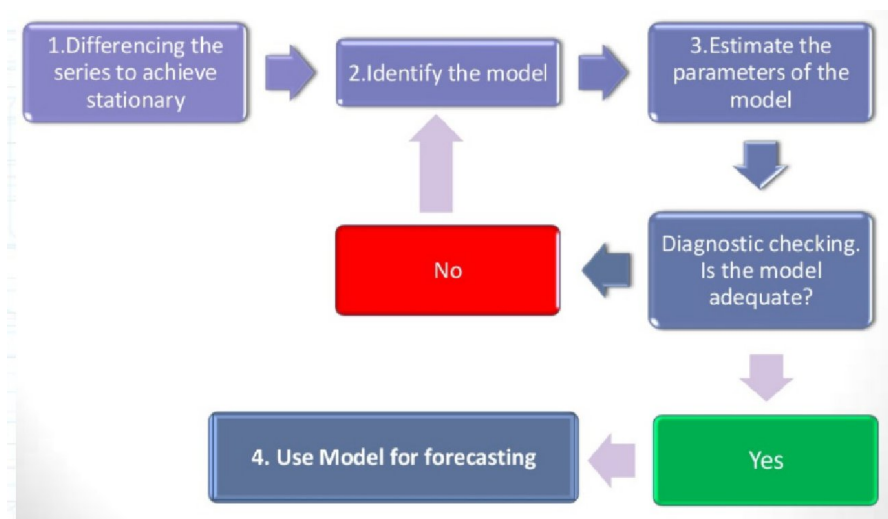


Рис. 1. Схематичне представлення підходу Бокса-Дженкінса до аналізу часових рядів (взято із роботи [5])

Таким чином, порядок різниці d визначається з умови стаціонарності різниць ряду $\Delta^d X_t$.

Пошук відповідних значень порядку моделі ARMA p та q в моделі ARMA(p, q) може бути полегшено шляхом побудови частинних автокореляційних функцій для оцінки p , а також використання автокореляційних функцій для оцінки q . Броквел та Девіс [15] для пошуку p і q радять застосовувати інформаційний критерій Акаїке (ІКА) [16].

Для моделі AR(p), якщо порядок задано p , то коефіцієнти моделі можна отримати за допомогою методу найменших квадратів, що призводить до відомої схеми Юла-Волкера [5], [13]:

$$a = \left(1 \ \rho_1 \ \rho_2 \ \rho_3 \ \dots \ \rho_{p-1} \ \rho_1 \ \rho_1 \ \rho_2 \ \dots \ \rho_{p-2} \ \rho_2 \ \rho_1 \ \rho_1 \ \dots \ \rho_{p-3} \ \rho_3 \ \rho_2 \ \rho_1 \ \dots \ \rho_{p-4} \ \dots \ \rho_{p-1} \ \rho_{p-2} \ \rho_{p-3} \ \rho_{p-4} \ \dots \ 1 \right)^{-1} \left(\rho_1 \ \rho_2 \ \rho_1 \ \dots \ \rho_{p-1} \ \rho_p \right), \quad (5)$$

де позначено ρ_i – автокореляції (значення індекса $i = 1, 2, \dots, p$), що розраховуються як відношення

$$\rho_i = \frac{R_i}{R_0} \quad (6)$$

і введено позначення для автоковаріації R_p , що розраховується з N спостережень ряду X_t :

$$R_i = \frac{1}{N-i} \sum_{t=1}^N X_t X_{t-i} \tag{7}$$

Параметри моделі MA(q) знаходяться методом найменших квадратів [5].

Покажемо тепер, як працюють моделі ARIMA в залежності від значення параметрів (p, d, q) на прикладі прогнозування чисельності населення України [17].

У даному випадку прийнято, що параметри приймають такі значення: p = 0..2, d = 0..1, q = 0..1, тобто всього порівнюється 12 моделей.

Результати знаходження параметрів моделей у системі програмування MATLAB 2019a [18] наведено нижче (див. табл. 1), параметри моделі розраховувались по всій сукупності даних за 1990-2019 рр. (зауважимо, що дані йдуть нерівномірно – 1990, 1995, 2000, 2001, 2002, ..., 2019) [17].

Таблиця 1.

Результати знаходження параметрів моделей ARIMA

1	ARIMA(0,0,0) Model (Gaussian Distribution):				
	Value	StandardError	TStatistic	PValue	
	Constant	46.155	0.60117	76.775	0
	Variance	7.6616	2.762	2.7739	0.0055393
2	ARIMA(0,0,1) Model (Gaussian Distribution):				
	Value	StandardError	TStatistic	PValue	
	Constant	46.575	0.76301	61.04	0
	MA{1}	0.82966	0.2623	3.163	0.0015615
	Variance	3.3174	1.0563	3.1405	0.0016866
3	ARIMA(0,1,0) Model (Gaussian Distribution):				
	Value	StandardError	TStatistic	PValue	
	Constant	-0.396	0.22958	-1.7249	0.08455
	Variance	0.42135	0.11943	3.5278	0.00041895
4	ARIMA(0,1,1) Model (Gaussian Distribution):				
	Value	StandardError	TStatistic	PValue	
	Constant	-0.40353	0.21686	-1.8608	0.062774
	MA{1}	-0.07561	0.78021	-0.09691	0.9228
	Variance	0.42009	0.1582	2.6554	0.0079221
5	ARIMA(1,0,0) Model (Gaussian Distribution):				
	Value	StandardError	TStatistic	PValue	
	Constant	0.40761	2.6958	0.1512	0.87982
	AR{1}	0.98296	0.052417	18.753	1.8363e-78
	Variance	0.41823	0.13091	3.1949	0.0013989
6	ARIMA(1,0,1) Model (Gaussian Distribution):				
	Value	StandardError	TStatistic	PValue	

	<table border="1"> <thead> <tr> <th></th> <th>Value</th> <th>StandardError</th> <th>TStatistic</th> <th>PValue</th> </tr> </thead> <tbody> <tr> <td>Constant</td> <td>0.48242</td> <td>2.3834</td> <td>0.20241</td> <td>0.8396</td> </tr> <tr> <td>AR{1}</td> <td>0.98111</td> <td>0.046646</td> <td>21.033</td> <td>3.2645e-98</td> </tr> <tr> <td>MA{1}</td> <td>-0.11491</td> <td>0.84306</td> <td>-0.1363</td> <td>0.89159</td> </tr> <tr> <td>Variance</td> <td>0.41581</td> <td>0.19311</td> <td>2.1533</td> <td>0.031295</td> </tr> </tbody> </table>		Value	StandardError	TStatistic	PValue	Constant	0.48242	2.3834	0.20241	0.8396	AR{1}	0.98111	0.046646	21.033	3.2645e-98	MA{1}	-0.11491	0.84306	-0.1363	0.89159	Variance	0.41581	0.19311	2.1533	0.031295					
	Value	StandardError	TStatistic	PValue																											
Constant	0.48242	2.3834	0.20241	0.8396																											
AR{1}	0.98111	0.046646	21.033	3.2645e-98																											
MA{1}	-0.11491	0.84306	-0.1363	0.89159																											
Variance	0.41581	0.19311	2.1533	0.031295																											
7	<p>ARIMA(1,1,0) Model (Gaussian Distribution):</p> <table border="1"> <thead> <tr> <th></th> <th>Value</th> <th>StandardError</th> <th>TStatistic</th> <th>PValue</th> </tr> </thead> <tbody> <tr> <td>Constant</td> <td>-0.37438</td> <td>0.28479</td> <td>-1.3146</td> <td>0.18865</td> </tr> <tr> <td>AR{1}</td> <td>0.077577</td> <td>0.37164</td> <td>0.20874</td> <td>0.83465</td> </tr> <tr> <td>Variance</td> <td>0.41821</td> <td>0.15455</td> <td>2.7059</td> <td>0.0068112</td> </tr> </tbody> </table>		Value	StandardError	TStatistic	PValue	Constant	-0.37438	0.28479	-1.3146	0.18865	AR{1}	0.077577	0.37164	0.20874	0.83465	Variance	0.41821	0.15455	2.7059	0.0068112										
	Value	StandardError	TStatistic	PValue																											
Constant	-0.37438	0.28479	-1.3146	0.18865																											
AR{1}	0.077577	0.37164	0.20874	0.83465																											
Variance	0.41821	0.15455	2.7059	0.0068112																											
8	<p>ARIMA(1,1,1) Model (Gaussian Distribution):</p> <table border="1"> <thead> <tr> <th></th> <th>Value</th> <th>StandardError</th> <th>TStatistic</th> <th>PValue</th> </tr> </thead> <tbody> <tr> <td>Constant</td> <td>-0.30691</td> <td>0.3324</td> <td>-0.92331</td> <td>0.35584</td> </tr> <tr> <td>AR{1}</td> <td>0.32956</td> <td>0.622</td> <td>0.52985</td> <td>0.59622</td> </tr> <tr> <td>MA{1}</td> <td>-0.47168</td> <td>0.45908</td> <td>-1.0274</td> <td>0.30421</td> </tr> <tr> <td>Variance</td> <td>0.39334</td> <td>0.21941</td> <td>1.7928</td> <td>0.073012</td> </tr> </tbody> </table>		Value	StandardError	TStatistic	PValue	Constant	-0.30691	0.3324	-0.92331	0.35584	AR{1}	0.32956	0.622	0.52985	0.59622	MA{1}	-0.47168	0.45908	-1.0274	0.30421	Variance	0.39334	0.21941	1.7928	0.073012					
	Value	StandardError	TStatistic	PValue																											
Constant	-0.30691	0.3324	-0.92331	0.35584																											
AR{1}	0.32956	0.622	0.52985	0.59622																											
MA{1}	-0.47168	0.45908	-1.0274	0.30421																											
Variance	0.39334	0.21941	1.7928	0.073012																											
9	<p>ARIMA(2,0,0) Model (Gaussian Distribution):</p> <table border="1"> <thead> <tr> <th></th> <th>Value</th> <th>StandardError</th> <th>TStatistic</th> <th>PValue</th> </tr> </thead> <tbody> <tr> <td>Constant</td> <td>0.97633</td> <td>2.7636</td> <td>0.35328</td> <td>0.72387</td> </tr> <tr> <td>AR{1}</td> <td>1.0962</td> <td>0.34473</td> <td>3.1799</td> <td>0.0014732</td> </tr> <tr> <td>AR{2}</td> <td>-0.12465</td> <td>0.33305</td> <td>-0.37429</td> <td>0.70819</td> </tr> <tr> <td>Variance</td> <td>0.4112</td> <td>0.16629</td> <td>2.4728</td> <td>0.013406</td> </tr> </tbody> </table>		Value	StandardError	TStatistic	PValue	Constant	0.97633	2.7636	0.35328	0.72387	AR{1}	1.0962	0.34473	3.1799	0.0014732	AR{2}	-0.12465	0.33305	-0.37429	0.70819	Variance	0.4112	0.16629	2.4728	0.013406					
	Value	StandardError	TStatistic	PValue																											
Constant	0.97633	2.7636	0.35328	0.72387																											
AR{1}	1.0962	0.34473	3.1799	0.0014732																											
AR{2}	-0.12465	0.33305	-0.37429	0.70819																											
Variance	0.4112	0.16629	2.4728	0.013406																											
10	<p>ARIMA(2,0,1) Model (Gaussian Distribution):</p> <table border="1"> <thead> <tr> <th></th> <th>Value</th> <th>StandardError</th> <th>TStatistic</th> <th>PValue</th> </tr> </thead> <tbody> <tr> <td>Constant</td> <td>-0.22804</td> <td>6.2558</td> <td>-0.036453</td> <td>0.97092</td> </tr> <tr> <td>AR{1}</td> <td>0.17142</td> <td>0.23515</td> <td>0.72897</td> <td>0.46602</td> </tr> <tr> <td>AR{2}</td> <td>0.81967</td> <td>0.17982</td> <td>4.5582</td> <td>5.16e-06</td> </tr> <tr> <td>MA{1}</td> <td>1</td> <td>0.37439</td> <td>2.671</td> <td>0.0075622</td> </tr> <tr> <td>Variance</td> <td>0.45008</td> <td>0.26921</td> <td>1.6719</td> <td>0.094548</td> </tr> </tbody> </table>		Value	StandardError	TStatistic	PValue	Constant	-0.22804	6.2558	-0.036453	0.97092	AR{1}	0.17142	0.23515	0.72897	0.46602	AR{2}	0.81967	0.17982	4.5582	5.16e-06	MA{1}	1	0.37439	2.671	0.0075622	Variance	0.45008	0.26921	1.6719	0.094548
	Value	StandardError	TStatistic	PValue																											
Constant	-0.22804	6.2558	-0.036453	0.97092																											
AR{1}	0.17142	0.23515	0.72897	0.46602																											
AR{2}	0.81967	0.17982	4.5582	5.16e-06																											
MA{1}	1	0.37439	2.671	0.0075622																											
Variance	0.45008	0.26921	1.6719	0.094548																											
11	<p>ARIMA(2,1,0) Model (Gaussian Distribution):</p> <table border="1"> <thead> <tr> <th></th> <th>Value</th> <th>StandardError</th> <th>TStatistic</th> <th>PValue</th> </tr> </thead> <tbody> <tr> <td>Constant</td> <td>-0.3678</td> <td>0.50055</td> <td>-0.7348</td> <td>0.46246</td> </tr> <tr> <td>AR{1}</td> <td>0.05824</td> <td>0.43384</td> <td>0.13424</td> <td>0.89321</td> </tr> <tr> <td>AR{2}</td> <td>0.056217</td> <td>0.34911</td> <td>0.16103</td> <td>0.87207</td> </tr> <tr> <td>Variance</td> <td>0.41694</td> <td>0.23346</td> <td>1.7859</td> <td>0.074113</td> </tr> </tbody> </table>		Value	StandardError	TStatistic	PValue	Constant	-0.3678	0.50055	-0.7348	0.46246	AR{1}	0.05824	0.43384	0.13424	0.89321	AR{2}	0.056217	0.34911	0.16103	0.87207	Variance	0.41694	0.23346	1.7859	0.074113					
	Value	StandardError	TStatistic	PValue																											
Constant	-0.3678	0.50055	-0.7348	0.46246																											
AR{1}	0.05824	0.43384	0.13424	0.89321																											
AR{2}	0.056217	0.34911	0.16103	0.87207																											
Variance	0.41694	0.23346	1.7859	0.074113																											
12	<p>ARIMA(2,1,1) Model (Gaussian Distribution):</p> <table border="1"> <thead> <tr> <th></th> <th>Value</th> <th>StandardError</th> <th>TStatistic</th> <th>PValue</th> </tr> </thead> <tbody> </tbody> </table>		Value	StandardError	TStatistic	PValue																									
	Value	StandardError	TStatistic	PValue																											

Constant	-0.28827	0.045926	-6.2768	3.4558e-10
AR{1}	0.75845	0.16442	4.613	3.9699e-06
AR{2}	-0.47021	0.095628	-4.9171	8.7843e-07
MA{1}	-1	0.28782	-3.4744	0.00051199
Variance	0.26513	0.11452	2.3151	0.020606

Результати прогнозування за допомогою моделей ARIMA наведено у табл. 2, прогноз розраховувався по попередніх значеннях у ковзаючому часовому вікні від 2003 по 2019 рр., 2020 рік – прогнозні значення, що розраховані по всьому діапазону моделювання (виділені у таблиці курсивом).

Для порівняння точності моделей було розраховано середньоквадратичне відхилення (СКВ) прогнозних і реальних даних (рис. 2а). Інший спосіб порівняння – розрахувати прогноз на 2020 рік і порівняти його з реальним значенням. Згідно даним Державної служби статистики України середня чисельність населення (за оцінкою) на кінець 2020 року становило 41.9 млн. осіб [18] (виділено у таблиці жирним курсивом). На рис. 2б показано різниці між прогнозними і реальними значеннями чисельності населення на 2020 р.

Таблиця 2.

Результати прогнозування за допомогою моделей ARIMA

p	Модель ARIMA												Реальні дані
	0	0	0	0	1	1	1	1	2	2	2	2	
	d	0	0	1	1	0	0	1	1	0	0	1	
q	0	1	0	1	0	1	0	1	0	1	0	1	
	1	2	3	4	5	6	7	8	9	10	11	12	
2003	49,9	49,1	47,7	47,8	47,7	47,5	47,6	47,2	49,4	49,2	47,1	48,0	47,8
2004	49,6	48,9	47,3	47,4	47,3	48,4	47,3	46,8	48,2	48,3	47,1	46,8	47,5
2005	49,3	48,8	47,1	47,1	47,1	47,1	47,0	46,5	47,7	47,7	47,6	47,1	47,1
2006	49,0	48,2	46,7	46,7	46,7	47,4	46,6	46,1	47,0	47,0	47,2	46,6	46,8
2007	48,7	48,2	46,4	46,4	46,4	46,4	46,4	45,8	46,7	46,8	46,3	46,3	46,5
2008	48,5	47,7	46,1	46,1	46,1	46,7	46,1	45,9	46,4	46,5	46,5	46,9	46,3
2009	48,3	47,7	45,9	46,0	46,0	46,1	45,9	45,8	46,2	46,3	46,3	46,3	46,1
2010	48,1	47,3	45,7	45,8	45,8	45,8	45,7	45,6	46,0	46,1	46,1	46,2	45,9
2011	48,0	47,3	45,5	45,6	45,6	45,7	45,5	45,5	45,8	45,9	45,9	45,6	45,7
2012	47,8	47,0	45,3	45,4	45,5	45,8	45,3	45,3	45,6	45,7	45,7	45,3	45,6
2013	47,7	47,0	45,3	45,3	45,4	45,5	45,3	45,2	45,6	45,6	45,6	45,0	45,5
2014	47,5	46,8	45,2	45,2	45,3	45,6	45,2	45,2	45,5	45,5	45,6	44,9	43,0
2015	47,3	44,7	42,6	42,9	42,6	40,6	42,1	42,7	42,1	41,6	42,4	42,1	42,8
2016	47,0	46,0	42,4	42,4	42,4	45,0	42,4	42,4	42,5	43,3	41,7	43,2	42,6
2017	46,8	44,5	42,2	42,1	42,2	42,2	42,2	42,1	42,3	42,9	42,2	42,7	42,4
2018	46,6	45,3	42,0	41,9	42,0	42,0	42,0	41,9	42,1	42,6	42,0	42,3	42,3
2019	46,4	44,4	41,9	41,9	42,0	42,0	41,9	41,8	42,1	42,3	41,9	42,0	42,0
2020	46,2	44,7	41,6	41,6	41,7	41,7	41,6	41,5	41,7	41,9	41,6	41,5	41,6
СКВ	1,1	0,8	0,6	0,6	0,6	1,1	0,6	0,6	0,7	0,7	0,7	0,6	
Відхилення від прогнозу	4,6	3,1	0,0	0,0	0,1	0,1	0,0	0,1	0,1	0,3	0,0	0,1	

Аналізуючи результати моделювання, бачимо, що у даному випадку найкращі результати за критерієм мінімуму середньоквадратичного відхилення дають моделі: **3** ($p=0, d=1, q=0$), **4** ($p=0, d=1, q=1$), **5** ($p=1, d=0, q=0$), **7** ($p=1, d=1, q=0$), **8** ($p=1, d=1, q=1$), **12** ($p=2, d=1, q=1$). Якщо ж розглядати критерій мінімального абсолютного відхилення від прогнозу на 2020 рік, то найкращими є моделі: **3** ($p=0, d=1, q=0$), **4** ($p=0, d=1, q=1$), **7** ($p=1, d=1, q=0$), **11** ($p=2, d=1, q=0$). Найкращою за сукупністю обох критеріїв є модель **7** ($p=1, d=1, q=0$). Таким чином, ми бачимо, вибір критерія оптимальності представляє собою окрему і складну задачу (хоча в даному випадку обидва критерії дають близькі результати). В даній роботі пропонується використовувати, як основний, критерій мінімуму СКВ, оскільки він дозволяє проводити селекцію моделей, спираючись тільки на наявні дані.

Нижче, на рис. 3, показано порівняння реальних і прогнозних значень для досліджуваних моделей ARIMA різних порядків.

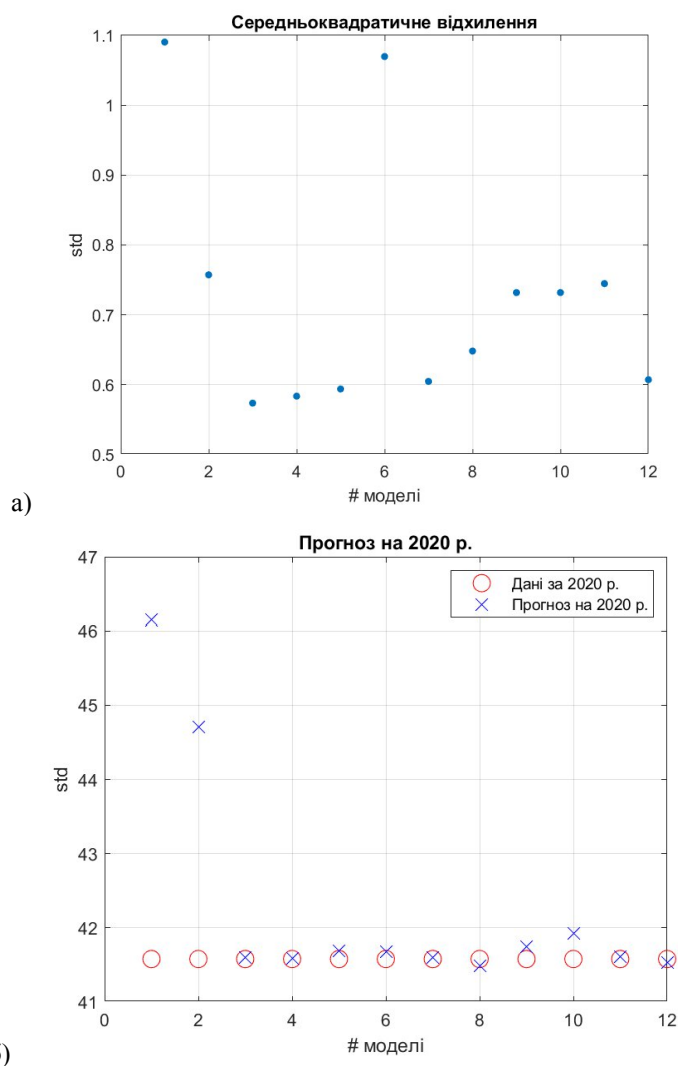
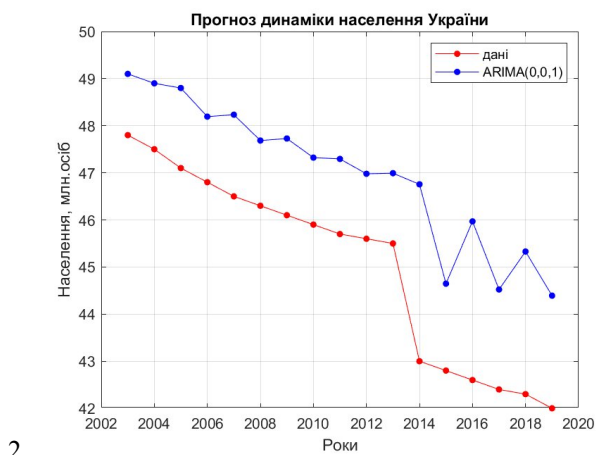
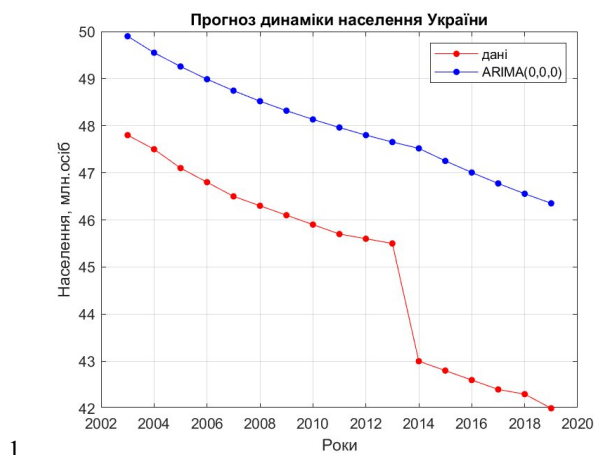


Рис. 2. Порівняння точності прогнозу: а) середньоквадратичні відхилення прогнозних і реальних значень; б) прогнозні і реальні значення на 2020 р.



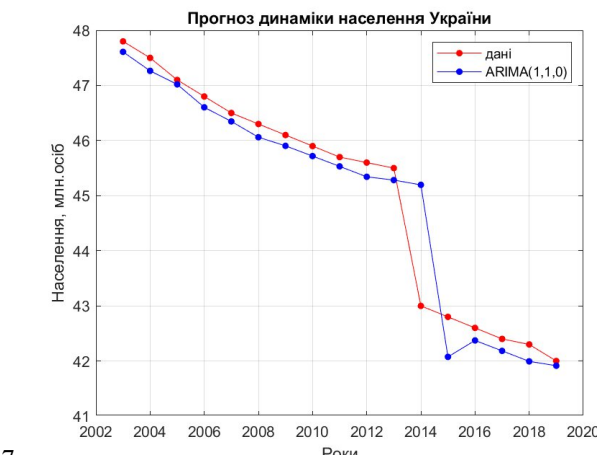
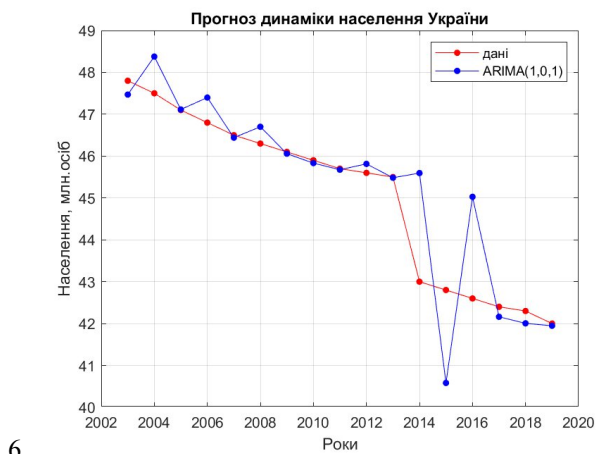
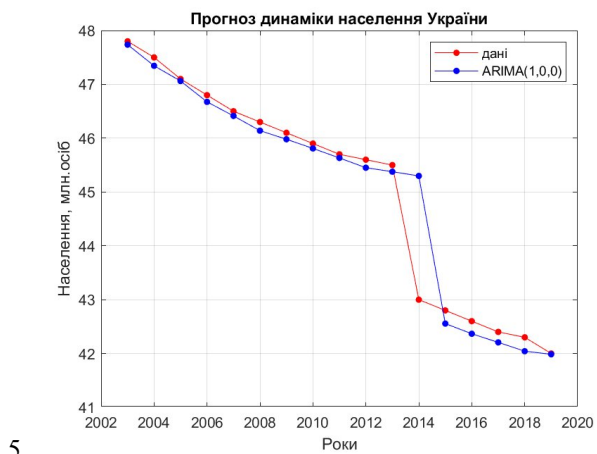
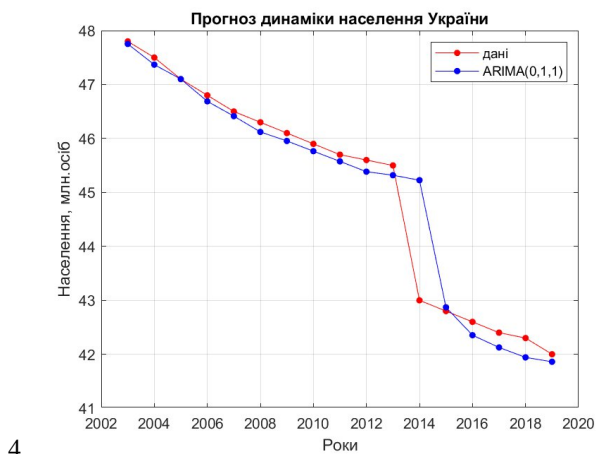
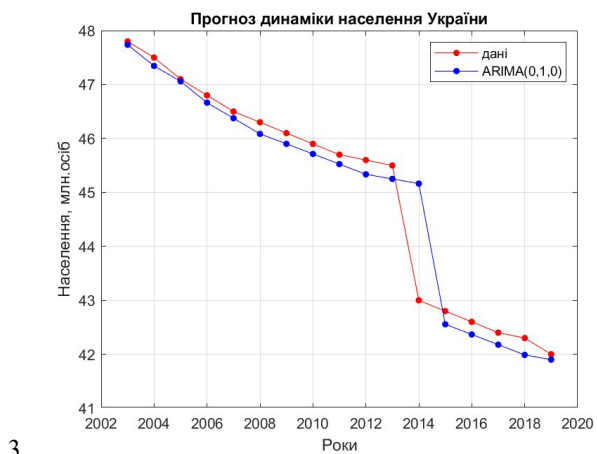


Рис. 3. Порівняння реальних і прогнозних значень для досліджуваних моделей ARIMA різних порядків

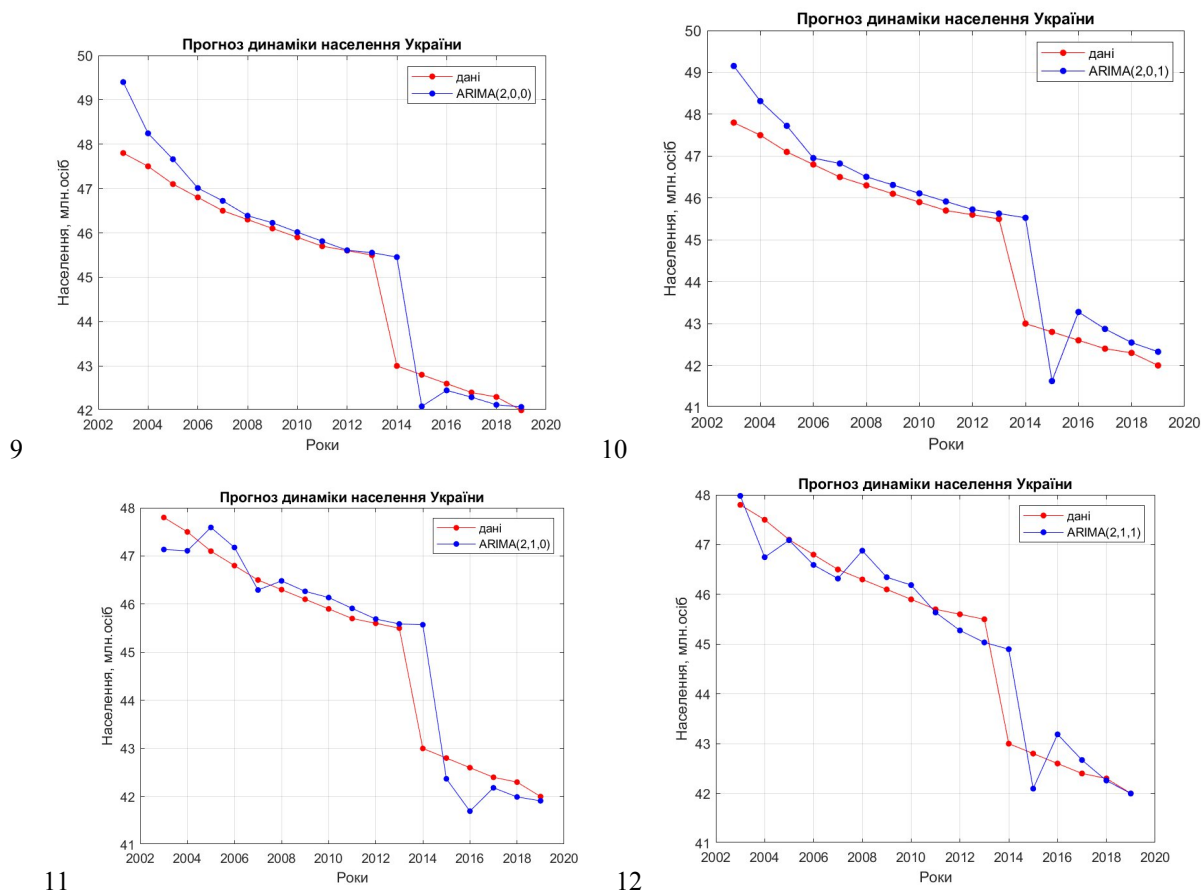


Рис. 3 (продовження). Порівняння реальних і прогнозних значень для досліджуваних моделей ARIMA різних порядків

Висновки та пропозиції.

У роботі проведено комплексне дослідження застосування ARIMA-моделей для прогнозування чисельності населення України за період 1990–2019 років. Показано, що моделі ARIMA з параметрами $(p=0, d=1, q=0)$, $(p=1, d=1, q=0)$ та $(p=2, d=1, q=0)$ забезпечують найкращі результати прогнозування за критеріями мінімального середньоквадратичного відхилення та абсолютної похибки прогнозу. Зроблено висновок, що для короткострокового прогнозування демографічних показників доцільно використовувати критерій мінімізації СКВ для вибору оптимальної моделі. Отримані результати можуть бути використані у системах підтримки прийняття рішень при стратегічному плануванні соціально-економічного розвитку країни.

Подальші дослідження рекомендується зосередити на застосуванні гібридних моделей ARIMA у поєднанні з методами машинного навчання для підвищення точності прогнозів у задачах з більш складною динамікою.

Декларація про штучний інтелект

Штучний інтелект не використовувався.

Конфлікт інтересів

Автори заявляють про відсутність конфлікту інтересів та підтверджують, що під час підготовки цієї роботи не існувало жодних комерційних, фінансових чи інших взаємовідносин, які могли б бути розцінені як такі, що здатні вплинути на результати дослідження або їх інтерпретацію. Робота виконана відповідно до принципів академічної доброчесності, етичних норм проведення наукових досліджень та вимог редакційної політики щодо запобігання конфлікту інтересів.

Список використаної літератури

1. Stiakakis, E., Katsimpras, G. (2023). *A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks*. *Future Internet*, 15(8), 255. DOI: 10.3390/fi15080255.
2. Zhang, G. P. (2003). *Time series forecasting using a hybrid ARIMA and neural network model*. *Neurocomputing*, 50, 159–175. DOI: 10.1016/S0925-2312(01)00702-0.

3. Guo, X. et al. (2021). *Applying an Improved Method Based on ARIMA Model to Predict the Short-Term Electricity Consumption Transmitted by the Internet of Things (IoT)*. Wireless Communications and Mobile Computing. DOI: 10.1155/2021/6610273.
4. Liu, Ü. Ç. B., Ertekin, Ş. (2018). *Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition*. arXiv preprint.
5. Box G. E. P., Jenkins G. M., Reinsel G. C. Time Series Analysis: Forecasting and Control. 3rd ed. Englewood Cliffs, NJ : Prentice Hall, 1994. 598 p.
6. Gan, Y., et al. (2010). *A hybrid ARIMA and SVM model for energy price forecasting*. In: Carbon Price Forecasting with a Hybrid ARIMA and Least Squares Support Vector Machines Methodology. DOI: 10.1007/978-3-319-57618-3_6.
7. Hyndman, R. J., Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. 3rd ed. OTexts.
8. Dong, Z., Zhou, Y. (2024). *A Novel Hybrid Model for Financial Forecasting Based on CEEMDAN-SE and ARIMA-CNN-LSTM*. Mathematics, 12(16), 2434. DOI: 10.3390/math12162434.
9. Next generation of GIS: must be easy / Z. A-Xing, Z. Fang-He, L. Peng, Q. Cheng-Zhi // Annals of GIS. – 2021. – № 27. – P. 71–86. DOI: 10.1080/19475683.2020.1766563.
10. Zhang, X., et al. (2023). *Hybrid Forecasting Model Combining ARIMA and LSTM for Time Series Prediction*. IEEE Access, 11, 55789–55800. DOI: 10.1109/ACCESS.2023.3267084.
11. Bandara, K., Bergmeir, C., Smyl, S. (2020). *Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach*. Expert Systems with Applications, 140, 112896. DOI: 10.1016/j.eswa.2019.112896.
12. Худов Г.В., Маковейчук О.М., Бутко І.М., Хижняк І.А. Модель прогнозування геопросторових даних в системах обробки геопросторової інформації. *Системи озброєння і військова техніка*. 2021. № 2(66). С. 123-128. <https://doi.org/10.30748/soivt.2021.66.16>.
13. Shumway R., Stoffer D. Time series analysis and its applications : with R examples. 3rd ed. New York : Springer, 2010. 596 p. ISBN 144197864X.
14. Enders W. *Stationary Time-Series Models // Applied Econometric Time Series*. 2nd ed. New York : Wiley, 2004. P. 48–107. ISBN 0-471-45173-8.
15. Brockwell P. J., Davis R. A. Time Series: Theory and Methods. 2nd ed. New York : Springer, 2009. 273 p. ISBN 9781441903198.
16. Akaike H. Information theory and an extension of the maximum likelihood principle // 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2–8, 1971. Budapest : Akadémiai Kiadó, 1973. P. 267–281. Reprinted in: Kotz S., Johnson N. L. (eds.) Breakthroughs in Statistics. Vol. I. New York : Springer-Verlag, 1992. P. 610–624.
17. Державна служба статистики України. Чисельність населення на 2020 рік [Електронний ресурс]. – Режим доступу: http://www.ukrstat.gov.ua/operativ/operativ2020/ds/kp/xls/kn1220_ue.xls.
18. MATLAB. Офіційний сайт [Електронний ресурс]. – Режим доступу: <https://www.mathworks.com/products/matlab.html>.

References

1. Stiakakis E., Katsimpras G. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. Future Internet. 2023. Vol. 15, No. 8. P. 255. DOI: 10.3390/fi15080255.
2. Zhang G. P. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 2003. Vol. 50. P. 159–175. DOI: 10.1016/S0925-2312(01)00702-0.
3. Guo X., et al. Applying an improved method based on ARIMA model to predict the short-term electricity consumption transmitted by the Internet of Things (IoT). Wireless Communications and Mobile Computing. 2021. Article ID 6610273. DOI: 10.1155/2021/6610273.
4. Liu Ü. Ç. B., Ertekin Ş. Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. arXiv preprint. 2018.
5. Box G. E. P., Jenkins G. M., Reinsel G. C. Time series analysis: forecasting and control. 3rd ed. Englewood Cliffs, NJ : Prentice Hall, 1994. 598 p.
6. Gan Y., et al. A hybrid ARIMA and SVM model for energy price forecasting. In: Carbon price forecasting with a hybrid ARIMA and least squares support vector machines methodology. 2010. DOI: 10.1007/978-3-319-57618-3_6.
7. Hyndman R. J., Athanasopoulos G. Forecasting: principles and practice. 3rd ed. OTexts, 2021.
8. Dong Z., Zhou Y. A novel hybrid model for financial forecasting based on CEEMDAN-SE and ARIMA-CNN-LSTM. Mathematics. 2024. Vol. 12, No. 16. P. 2434. DOI: 10.3390/math12162434.
9. Xing Z. A., Fang-He Z., Peng L., Cheng-Zhi Q. Next generation of GIS: must be easy. Annals of GIS. 2021. Vol. 27. P. 71–86. DOI: 10.1080/19475683.2020.1766563.
10. Zhang X., et al. Hybrid forecasting model combining ARIMA and LSTM for time series prediction. IEEE Access. 2023. Vol. 11. P. 55789–55800. DOI: 10.1109/ACCESS.2023.3267084.
11. Bandara K., Bergmeir C., Smyl S. Forecasting across time series databases using recurrent neural networks on groups of similar series: a clustering approach. Expert Systems with Applications. 2020. Vol. 140. Article 112896. DOI: 10.1016/j.eswa.2019.112896.

12. Khudov H. V., Makoveychuk O. M., Butko I. M., Khizhnyak I. A. Model of geospatial data forecasting in geoinformation processing systems. *Armament and Military Equipment Systems*. 2021. No. 2(66). P. 123–128. DOI: 10.30748/soivt.2021.66.16.

13. Shumway R., Stoffer D. *Time series analysis and its applications: with R examples*. 3rd ed. New York : Springer, 2010. 596 p. ISBN 144197864X.

14. Enders W. Stationary time-series models. In: *Applied econometric time series*. 2nd ed. New York : Wiley, 2004. P. 48–107. ISBN 0-471-45173-8.

15. Brockwell P. J., Davis R. A. *Time series: theory and methods*. 2nd ed. New York : Springer, 2009. 273 p. ISBN 9781441903198.

16. Akaike H. Information theory and an extension of the maximum likelihood principle. In: *2nd International Symposium on Information Theory*, Tsahkadsor, Armenia, USSR, 1971. Budapest : Akadémiai Kiadó, 1973. P. 267–281. Reprinted in: Kotz S., Johnson N. L., eds. *Breakthroughs in Statistics*. Vol. 1. New York : Springer-Verlag, 1992. P. 610–624.

17. State Statistics Service of Ukraine. Population data for 2020 [Electronic resource]. Available at: http://www.ukrstat.gov.ua/operativ/operativ2020/ds/kn/xls/kn1220_ue.xls.

18. MATLAB official website [Electronic resource]. Available at: <https://www.mathworks.com/products/matlab.html>.

Надійшла до редакції: 20.11.25

Прийнята до друку: 12.06.26

Опубліковано: 30.06.26