

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ  
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС  
«ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ»  
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

## СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

МІЖНАРОДНИЙ НАУКОВО-ТЕХНІЧНИЙ ЖУРНАЛ

№ 3

2022

ЗАСНОВАНО У ЛИПНІ 2001 р.

**РЕДАКЦІЙНА КОЛЕГІЯ:**

**Головний редактор**

**М.З. ЗГУРОВСЬКИЙ**, акад. НАН України

**Заступник головного редактора**

**Н.Д. ПАНКРАТОВА**, чл.-кор. НАН України

**Члени редколегії:**

**П.І. АНДОН**, акад. НАН України

**А.В. АНІСІМОВ**, чл.-кор. НАН України

**Х. ВАЛЕРО**, проф., Іспанія

**Г.-В. ВЕБЕР**, проф., Турція

**П.О. КАСЬЯНОВ**, проф., д.ф.-м.н.,  
Україна

**Й. КОРБИЧ**, проф. Польща

**О.А. ПАВЛОВ**, проф., д.т.н., Україна

**Л. САКАЛАУСКАС**, проф., Литва

**А.М. САЛЕМ**, проф., Єгипет

**І.В. СЕРГІЄНКО**, акад. НАН України

**Х.-М. ТЕОДОРЕСКУ**, акад. Румунської  
Академії

**Е.О. ФАЙНБЕРГ**, проф., США

**Я.С. ЯЦКІВ**, акад. НАН України

У номері:

• **Теоретичні та прикладні проблеми і методи системного аналізу**

• **Прогресивні інформаційні технології, високопродуктивні комп'ютерні системи**

• **Проблеми прийняття рішень та управління в економічних, технічних, екологічних і соціальних системах**

• **Теоретичні та прикладні проблеми інтелектуальних систем підтримання прийняття рішень**

**АДРЕСА РЕДАКЦІЇ:**

03056, м. Київ,

просп. Перемоги, 37, корп. 35,

ННК «ІПСА» КПІ ім. Ігоря Сікорського

Тел.: 204-81-44; факс: 204-81-44

E-mail: journal.iasa@gmail.com

http://journal.iasa.kpi.ua

NATIONAL ACADEMY OF SCIENCES OF UKRAINE  
EDUCATIONAL AND SCIENTIFIC COMPLEX  
«INSTITUTE FOR APPLIED SYSTEM ANALYSIS»  
OF THE NATIONAL TECHNICAL UNIVERSITY OF UKRAINE  
«IGOR SIKORSKY KYIV POLYTECHNIC INSTITUTE»

## SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

INTERNATIONAL SCIENTIFIC AND TECHNICAL JOURNAL

№ 3

2022

IT IS FOUNDED IN JULY 2001

### EDITORIAL BOARD:

#### The editor – in – chief

**M.Z. ZGUROVSKY,** Academician of  
NASU

#### Deputy editor – in – chief

**N.D. PANKRATOVA,** Correspondent  
member of NASU

#### Associate editors:

**F.I. ANDON,** Academician of  
NASU

**A.V. ANISIMOV,** Correspondent  
member of NASU

**E.A. FEINBERG,** Prof., USA

**P.O. KASYANOV,** Prof., Ukraine

**J. KORBICH,** Prof., Poland

**A.A. PAVLOV,** Prof., Ukraine

**L. SAKALAIUSKAS,** Prof., Lithuania

**A.M. SALEM,** Prof., Egypt

**I.V. SERGIENKO,** Academician of NASU

**H.-N. TEODORESCU,** Academician of  
Romanian Academy

**J. VALERO** Prof., Spain

**G.-W. WEBER,** Prof., Turkey

**Ya.S. YATSKIV,** Academician of NASU

### THE EDITION ADDRESS:

03056, Kyiv,  
av. Peremogy, 37, building 35,  
Institute for Applied System Analysis  
at the Igor Sikorsky Kyiv Polytechnic Institute  
Phone: **204-81-44**; Fax: **204-81-44**  
E-mail: [journal.iasa@gmail.com](mailto:journal.iasa@gmail.com)  
<http://journal.iasa.kpi.ua>

### In the issue:

- **Theoretical and applied problems and methods of system analysis**
- **Progressive information technologies, high-efficiency computer systems**
- **Decision making and control in economic, technical, ecological and social systems**
- **Theoretical and applied problems of intelligent systems for decision making support**

## Шановні читачі!

Навчально-науковий комплекс «Інститут прикладного системного аналізу» Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» видає міжнародний науково-технічний журнал

### «СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ».

Журнал публікує праці теоретичного та прикладного характеру в широкому спектрі проблем, що стосуються системних досліджень та інформаційних технологій.

#### Провідні тематичні розділи журналу:

Теоретичні та прикладні проблеми і методи системного аналізу; теоретичні та прикладні проблеми інформатики; автоматизовані системи управління; прогресивні інформаційні технології, високопродуктивні комп'ютерні системи; проблеми прийняття рішень і управління в економічних, технічних, екологічних і соціальних системах; теоретичні та прикладні проблеми інтелектуальних систем підтримання прийняття рішень; проблемно і функціонально орієнтовані комп'ютерні системи та мережі; методи оптимізації, оптимальне управління і теорія ігор; математичні методи, моделі, проблеми і технології дослідження складних систем; методи аналізу та управління системами в умовах ризику і невизначеності; евристичні методи та алгоритми в системному аналізі та управлінні; нові методи в системному аналізі, інформатиці та теорії прийняття рішень; науково-методичні проблеми в освіті.

**Головний редактор журналу** — ректор Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», академік НАН України Михайло Захарович Згуровський.

Журнал «Системні дослідження та інформаційні технології» включено до переліку фахових видань ВАК України.

Журнал «Системні дослідження та інформаційні технології» входить до таких наукометричних баз даних: Scopus, EBSCO, Google Scholar, DOAJ, Index Copernicus, реферативна база даних «Україніка наукова», український реферативний журнал «Джерело», наукова періодика України.

Статті публікуються українською та англійською мовами.

Журнал рекомендовано передплатити. **Наш індекс 23918.** Якщо ви не встигли передплатити журнал, його можна придбати безпосередньо в редакції за адресою: 03056, м. Київ, просп. Перемоги, 37, корп. 35.

Завідувачка редакції **С.М. Шевченко**

Редакторка **Р.М. Шульженко**

Молодша редакторка **Л.О. Тарин**

Комп'ютерна верстка, дизайн **А.А. Патіюхи**

Свідоцтво про реєстрацію КВ № 23234–13074 ПР від 22.03.2018 р.

---

Підписано до друку 30.10.2022. Формат 70x108 1/16. Папір офс. Гарнітура Times.

Спосіб друку – цифровий. Ум. друк. арк. 14,411. Обл.-вид. арк. 28,56. Наклад 100 пр. Зам. № 11/04

---

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Свідоцтво про державну реєстрацію: ДК № 5354 від 25.05.2017 р.

просп. Перемоги, 37, м. Київ, 03056.

ФОП Пилипенко Н.М., вул. Мічуріна, б. 2/7, м. Київ, 01014.

Виписка з Єдиного державного реєстру № 2 070 000 0000 0214697 від 17.05.2019 р.,

тел. (044) 361 78 68.

## **Dear Readers!**

Educational and Scientific Complex «Institute for Applied System Analysis» of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» is published of the international scientific and technical journal

### **«SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES».**

The Journal is printing works of a theoretical and applied character on a wide spectrum of problems, connected with system researches and information technologies.

#### **The main thematic sections of the Journal are the following:**

Theoretical and applied problems and methods of system analysis; theoretical and applied problems of computer science; automated control systems; progressive information technologies, high-efficiency computer systems; decision making and control in economic, technical, ecological and social systems; theoretical and applied problems of intellectual systems for decision making support; problem- and function-oriented computer systems and networks; methods of optimization, optimum control and theory of games; mathematical methods, models, problems and technologies for complex systems research; methods of system analysis and control in conditions of risk and uncertainty; heuristic methods and algorithms in system analysis and control; new methods in system analysis, computer science and theory of decision making; scientific and methodical problems in education.

**The editor-in-chief of the Journal** is rector of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», academician of the NASU Michael Zaharovich Zgurovsky.

The articles to be published in the Journal in Ukrainian and English languages are accepted. Information printed in the Journal is included in the Catalogue of periodicals of Ukraine.

# СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

3 • 2022

## ЗМІСТ

<b>ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ І МЕТОДИ СИСТЕМНОГО АНАЛІЗУ</b>	
<i>Zgurovsky M., Pyshnograiev I.</i> Study of security trends of the global society based on intelligent data analysis .....	7
<b>ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ВИСОКОПРОДУКТИВНІ КОМП'ЮТЕРНІ СИСТЕМИ</b>	
<i>Yatsenko V., Kruchinin S., Bidyuk P.</i> Superconducting gravimeters based on advanced nanomaterials and quantum neural network .....	30
<b>ПРОБЛЕМИ ПРИЙНЯТТЯ РІШЕНЬ ТА УПРАВЛІННЯ В ЕКОНОМІЧНИХ, ТЕХНІЧНИХ, ЕКОЛОГІЧНИХ І СОЦІАЛЬНИХ СИСТЕМАХ</b>	
<i>Romanenko V.D., Miliavskiy Y.L.</i> Combined control of multirate impulse processes in a cognitive map of COVID-19 morbidity .....	46
<i>Kovalets I.V., Besspalov V.P., Maistrenko S.Ya., Udovenko O.I.</i> The use of environmental decision support systems for modeling of atmospheric pollution following the chemical accidents .....	57
<i>Астраханцев А.А., Ляшенко Г.Є.</i> Процес керування захищеністю даних під час віддаленої біометричної автентифікації .....	71
<b>ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМАННЯ ПРИЙНЯТТЯ РІШЕНЬ</b>	
<i>Vijayasekaran G., Duraipandian M.</i> Resource scheduling in edge computing IoT networks using hybrid deep learning algorithm .....	86
<i>Федін С.С., Зубрецька І.С.</i> Методи та моделі нейромережевої апроксимації градувальних характеристик NTC-термісторів .....	102
<i>Sandeep Kumar Bothra, Sunita Singhal, Hemlata Goyal.</i> Cost effective hybrid genetic algorithm for workflow scheduling in cloud .....	121
<i>Тутаренко А.</i> Multi-step prediction in linearized latent state spaces for representation learning .....	139
Відомості про авторів .....	149

# SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

3 • 2022

## CONTENT

<b>THEORETICAL AND APPLIED PROBLEMS AND METHODS OF SYSTEM ANALYSIS</b>	
<i>Zgurovsky M., Pyshnograiev I.</i> Study of security trends of the global society based on intelligent data analysis .....	7
<b>PROGRESSIVE INFORMATION TECHNOLOGIES, HIGH-EFFICIENCY COMPUTER SYSTEMS</b>	
<i>Yatsenko V., Kruchinin S., Bidyuk P.</i> Superconducting gravimeters based on advanced nanomaterials and quantum neural network .....	30
<b>DECISION MAKING AND CONTROL IN ECONOMIC, TECHNICAL, ECOLOGICAL AND SOCIAL SYSTEMS</b>	
<i>Romanenko V.D., Miliavskiy Y.L.</i> Combined control of multirate impulse processes in a cognitive map of COVID-19 morbidity .....	46
<i>Kovalets I.V., Bepalov V.P., Maistrenko S.Ya., Udovenko O.I.</i> The use of environmental decision support systems for modeling of atmospheric pollution following the chemical accidents .....	57
<i>Astrakhantsev A.A., Liashenko G.Ye.</i> Data protection management process during remote biometric authentication .....	71
<b>THEORETICAL AND APPLIED PROBLEMS OF INTELLIGENT SYSTEMS FOR DECISION MAKING SUPPORT</b>	
<i>Vijayasekaran G., Duraipandian M.</i> Resource scheduling in edge computing IoT networks using hybrid deep learning algorithm .....	86
<i>Fedin S.S., Zubretska I.S.</i> Methods and models of neural networks for approximation of calibration characteristics of NTC-thermistors .....	102
<i>Sandeep Kumar Bothra, Sunita Singhal, Hemlata Goyal.</i> Cost effective hybrid genetic algorithm for workflow scheduling in cloud .....	121
<i>Tytarenko A.</i> Multi-step prediction in linearized latent state spaces for representation learning .....	139
Information about the authors .....	149

## STUDY OF SECURITY TRENDS OF THE GLOBAL SOCIETY BASED ON INTELLIGENT DATA ANALYSIS

M. ZGUROVSKY, I. PYSHNOGRAIEV

**Abstract.** This article is devoted to applying system analysis and data mining methodology to one of the most pressing problems today: studying the security of a global society in a conflicting world. A set of global threats relevant to the first half of the 21st century is considered. These threats have been identified by the United Nations (UN), the World Health Organization (WHO), the World Economic Forum, and other reputable international organizations. As a result of applying the Delphi method to analyze a wide range of threats identified by these organizations, 11 of the most important threats to humanity in the first half of the 21st century were identified. The vulnerabilities of different countries to the impact of the totality of these threats are analyzed. Scenarios for the possible development of a global society during and after the conflict are constructed.

**Keywords:** global safety, systemic conflicts, global threats, Minkowski norm, vulnerability.

### INTRODUCTION

Since the beginning of the 21st century, many recognized international organizations have conducted research to identify the major challenges facing humanity. Such organizations include the United Nations (UN), the World Health Organization (WHO), the World Economic Forum (WEF), Transparency International, the Global Footprint Network, the International Energy Agency, the World Resources Institute, the British Petroleum Company and others. Each of these organizations not only identified challenges for their field of activity, but also tried to assess the impact of these challenges on other areas of human life.

There is a problem of consolidating these studies and creating a necessary and sufficient set of global threats to the sustainable development of mankind.

This study is a continuation of studies of the behaviour of complex socio-economic systems [1], global threats and sustainable development processes [2]. The new study took into account the results of the analysis of global threats to humanity, performed by the following international organizations:

1. On January 11, 2022, the World Economic Forum presented The Global Risks Report 2022 [3], in which for the next 10 years it formed the necessary and sufficient set of threats to the sustainable development of mankind. WEF experts

identified a total of 37 global threats in 5 areas of human activity: economic, environmental, geopolitical, social and technological.

2. Using the Delphi method, The Millennium Project identified 15 global challenges in the same areas [4].

3. Due to the fact that this study examines the threats to sustainable development, it is also necessary to take into account the 17 UN Sustainable Development Goals set out in the “Sustainable Development Agenda 2030” [5, 6].

For further study, we use the variety of threats formulated by the above-mentioned international organizations.

## CHARACTERISTICS OF GLOBAL THREATS TO SUSTAINABLE DEVELOPMENT

As a result of applying the Delphi method to analyze a wide range of threats identified by the organizations mentioned above, 11 of the most important threats to humanity in the first half of the 21st century were identified.

### Threat 1. Global decrease in energy security (ES)

The country’s energy independence is an integral and fundamental component of its sovereignty. It determines the country’s self-sufficiency in energy supply and energy generation. In the conditions of constant growth of consumption of energy of the world it is necessary to increase also its production (Fig. 1).

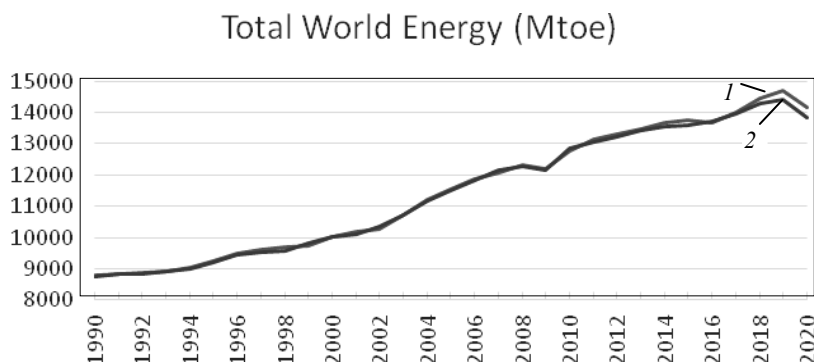


Fig. 1. World energy consumption and production: 1 — Production; 2 — Consumption

Source: based on data in [7]

At the same time, it is impossible to constantly increase the extraction of fossil energy resources (Fig. 2), whose reserves are rapidly declining. In addition, the behaviour of this type of resources in the market is significantly influenced by world politics. In the Short-term Energy Outlook of the U.S. The Energy Information Administration [8] noted that the oil and gas market have great uncertainties, including due to “Russia’s full-scale invasion of Ukraine”.

On the one hand, the world community needs to find new energy sources, develop alternative energy [9], and on the other hand in conditional of changing geopolitics, the is a need to take care of its independence from extremal energy supplies.



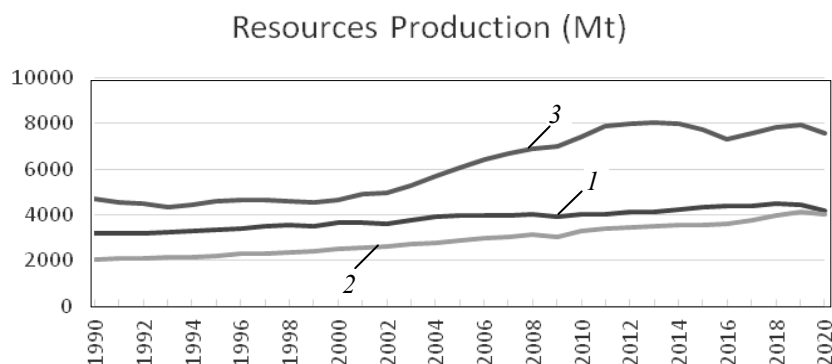


Fig. 2. Production of fossil resources: 1 — Crude oil; 2 — Natural gas; 3 — Coal and lignite

Source: based on data in [7]

In order to quantitatively estimate the energy security of different countries of the world the Energy Freedom Index (ES) [10] is used. It aggregates the of three separate sub-indices, which can be the object of independent analysis:

- Sub-index of energy potential – determines the established potential of the country in terms of access to fuel and energy resources: coal, natural gas and crude oil reserves (calculated as the value of the total explored reserves of coal, natural gas and crude oil, determined per capita).
- Sub-index of energy balance – reflects the annual balance between total production and consumption of electricity and heat in the country (calculated as the ratio of annual production and annual energy consumption in million metric tons of oil equivalent).
- Sub-index of energy development – demonstrates the ability of the country's energy system to develop with the possibility of energy transition (calculated as chain growth rate of the total installed capacity of all electricity generation facilities in the country).

## Threat 2. The imbalance between biological capacity of the Earth and human needs in biosphere (BB)

In early 2022, the world's population reached 7.95 billion people living on the total area 510 072 000 km<sup>2</sup> [11]. According to the method of arithmetic extrapolation the Earth population will have been 9.75 billion people by the year 2050. At the same time, our planet has limited space and resources.

In 2018, the consumption of natural resources exceeded 1.75 times that the Earth's biosphere can restore, forming a significant environmental deficit (Fig. 3).

Ecological Footprint adds up all the productive areas for which a population, a person or a product competes. It measures the ecological assets that a given population or product requires to produce the natural resources it consumes (including plant-based food and fiber products, livestock and fish products, timber and other forest products, space for urban infrastructure) and to absorb its waste, especially carbon emissions. The Ecological Footprint tracks the use of productive surface areas. Typically, these areas are: cropland, grazing land, fishing grounds, built-up land, forest area, and carbon demand on land. On the supply side, a city, state or nation's biocapacity represents the productivity of its ecologi-

cal assets (including cropland, grazing land, forest land, fishing grounds, and built-up land) [13].

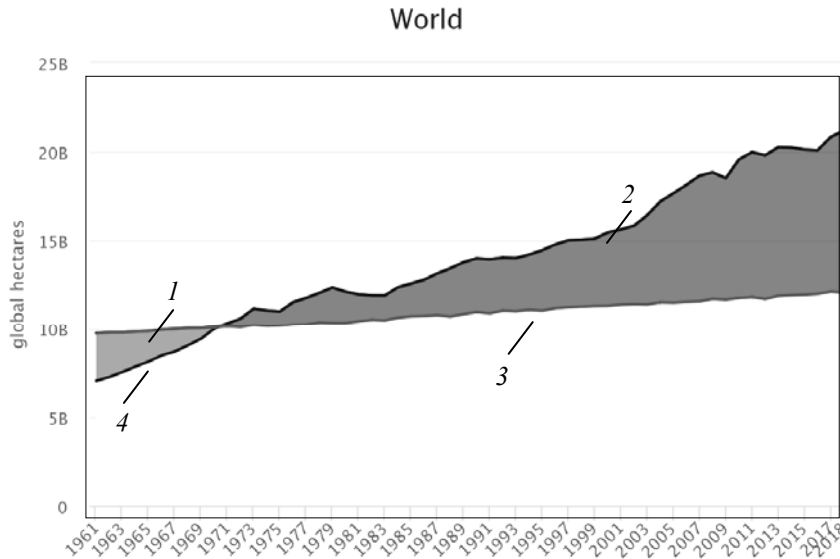


Fig. 3. Ecological footprint trend: 1 — Ecological Reserve; 2 — Ecological Deficit; 3 — Biocapacity; 4 — Ecological Footprint

Source: [12]

Therefore, the increase in the ecological deficit over time can lead to irreversible changes in the biosphere, which will directly threaten the existence of mankind.

For estimation of increasing threats, connected with imbalance between biological capability of the Earth and human requirements in biosphere, in terms of demographic structure change of the world we will use the indicator which is ratio level between biocapacity and ecological footprint consumption for a country [12]:

- value  $>1$  – the country is an ecological creditor;
- value  $<1$  – the country is an ecological debtor.

### Threat 3. Growing inequality between people and countries on the Earth (GINI)

According to the World Bank, in 2018, 3 billion people live on less than \$ 150 a month [14]. And although most regions, except the Middle East and North Africa, are showing progress in the fight against poverty, the situation remains threatening [15].

Political and military conflicts, pandemics, global corruption, terrorism, depletion of resources, etc. complicate humanity's ability to overcome poverty and inequality. For example, [16] emphasizes that due to the restrictions imposed by the proliferation of Covid-19, for the first time since 1993, inequalities between countries are projected to increase (Fig. 4).

To assess quantitatively the disparity of the distribution of economic and social benefits for each of the countries under study, we will use the Gini index [17], which reflects these characteristics.

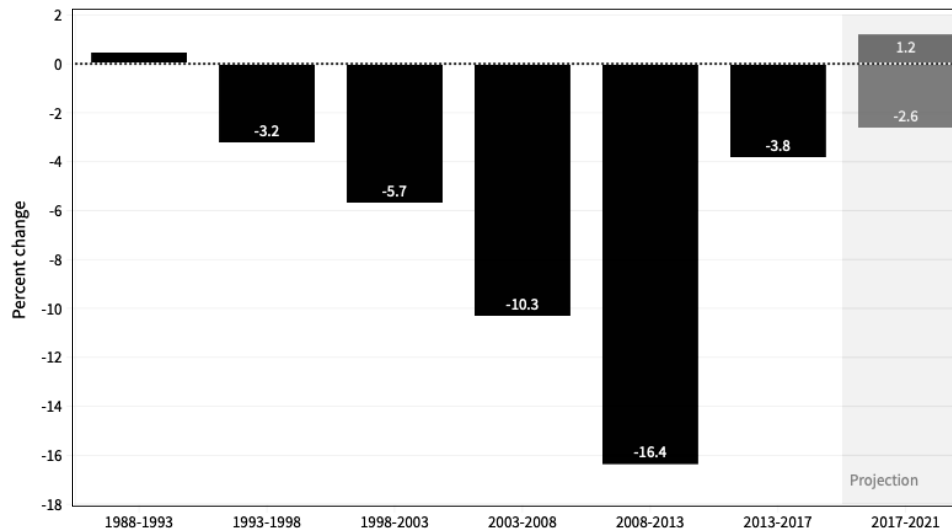


Fig. 4. Degree of inequality between countries

Source: [16]

#### Threat 4. The spread of global diseases (GD)

The World Health Organization has identified the top 10 causes of death globally in 2019, which caused 55% of 55.4 million deaths worldwide (Fig. 5) [18].

At a global level, 7 of the 10 leading causes of deaths in 2019 were non-communicable diseases. They kill 41 million people each year, equivalent to 71% of all deaths globally. The main types of NCD are cardiovascular diseases (such as heart attacks and stroke), cancers, chronic respiratory diseases (such as chronic obstructive pulmonary disease and asthma) and diabetes [19].

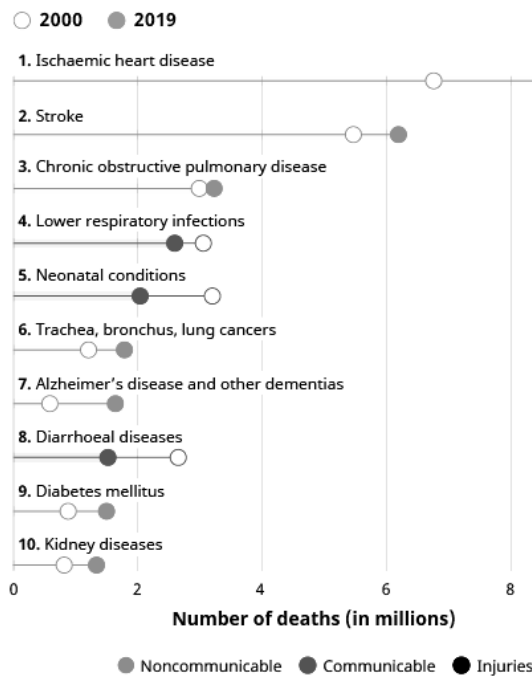


Fig. 5. Top 10 causes of death globally in 2019.

Source: [18]

Also, in the study we consider the communicable (including infectious) diseases, such as: tuberculosis, HIV/AIDS, diarrhea, malaria, hepatitis, etc. They have lower part of deaths globally, but the threat of new diseases remains. Lower respiratory infections remained the world's most deadly communicable disease, ranked as the 4th leading cause of death. However, the number of deaths has gone down substantially: in 2019 it claimed 2.6 million lives, 460 000 fewer than in 2000 [20]. Also, nearly half of the world's population was at risk of malaria in 2020, it is estimated 41 million cases [18]. In 2020, 680 000 (480 000–1.0 million) people died from HIV-related causes and 1.5 million (1.0–2.0 million) people acquired HIV [21]. Also dangerous for the world community are the pandemics of swine flu (2008–2009), Ebola (2014–2015), SARS-CoV-2 (from 2020), which also contributed to the deepening of economic crises [22].

The spread of global diseases (GD) is measured in the normalized total number of people (millions per year) who died from these diseases. For the subsequent simulation, we take data on these diseases from the World Health Organization [23].

### Threat 5. Information gap (IG)

Humanity is constantly generating gigantic volumes of new data and information. There were 79 zettabytes of data generated worldwide in 2021, 90% of it is replicated [24] (Fig. 6). This raises a number of challenges: how to access this information, how to process it, and whether it is trustworthy.

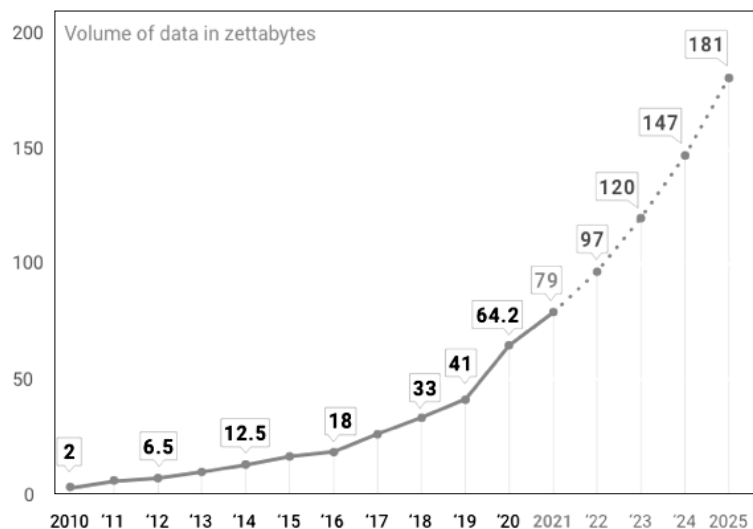


Fig. 6. Annual volume of data generated, consumed, copied and stored  
Source: [24]

To assess these challenges Information Gap is formed by following determinants of the modern information society:

1. Readiness of the local ICT infrastructure (RLI). This indicator is based on the ICT Development Index [25], it shows the degree of involvement of the population in the consumption / generation of information and the level of exports of services and goods of the sector, which indicates the availability of an appropriate base. The following data sets with equal weights are used for this purpose:

- individuals using the Internet (% of population) [26];
- mobile cellular subscriptions (per 100 people) [26];
- fixed broadband subscriptions (per 100 people) [26];
- ICT goods exports (% of total goods exports) [27];
- ICT service exports (BoP, current US\$) [28].

2. Number of secure Internet servers (per 1 million people) (SIS) [29]. Secure servers are servers using encryption technology in internet transactions. They provide the infrastructure for the secure exchange of generated data.

3. The vulnerability of one or another country, territory or world to the action of cyber-attacks. This component will be measured using the Global Cybersecurity Index (GCI) [30].

4. World Press Freedom index (WPF). It is defined as the ability of journalists as individuals and collectives to select, produce, and disseminate news in the public interest independent of political, economic, legal, and social interference and in the absence of threats to their physical and mental safety [31].

The resulting index of IG will be calculated from the formula:

$$IG = 0.4RLI + 0.15SIS + 0.3GCI + 0.15WPF.$$

#### **Threat 6. Corruption perception (CP)**

Humanity is constantly generating gigantic volumes of new data and information. There were 79 zettabytes of data generated worldwide in 2021, 90% of it is replicated [24] (Fig. 6). This raises a number of challenges: how to access this information, how to process it, and whether it is trustworthy.

Corruption is the biggest obstacle to the economic and social development of society. Over last decade world has made no significant progress against corruption [32].

To estimate the influence of corruption on socio-economical and cultural development of different countries of the world we will use the Corruption Perception Index established by the international organization Transparency International [33].

Corruption is connected with all spheres of society. Countries experiencing armed conflict or authoritarianism tend to earn the lowest scores, including Venezuela, Afghanistan, North Korea, Yemen, Equatorial Guinea, Libya and Turkmenistan. Also, last research of Transparency International showed that corruption level is opposite to the level of human rights [32].

#### **Threat 7. Limited access to drinking water (WA)**

According to the data of the World Health Organization (WHO) and the UNICEF the world is under the threat of reduced the access to drinking (potable) water and to sanitary facilities. The fourth part of all mankind (2 billion people) does not have access to drinking water in 2020. At the same time 46% (4.2 billion people) lack safe sanitation. This situation persists, provided that in 2030 the Agenda for Sustainable Development agreed to take concrete steps to achieve goals 6.1 and 6.2, i.e. to make access to water “for all” [34].

Meanwhile the world’s population grows, especially in underdeveloped countries, the struggle for control over the remnants of drinking water resources

increases. This phenomenon gives rise to the next, growing in time, threat to humanity.

The limited access to the drinking water will be estimated by the inversed magnitude to the indicator of the access to drinking water [35].

### Threat 8. Impact of climate change and natural disaster (CN)

According to [5], the threat of climate change and the occurrence of natural disasters require increased attention of society and its consolidated efforts to minimize this factor.

Since the 1940s, the Earth's surface temperature has been constantly rising (Fig. 7). That extra heat is driving regional and seasonal temperature extremes, reducing snow cover and sea ice, intensifying heavy rainfall, increase the number of natural disasters, and changing habitat ranges for plants and animals – expanding some and shrinking others [36].

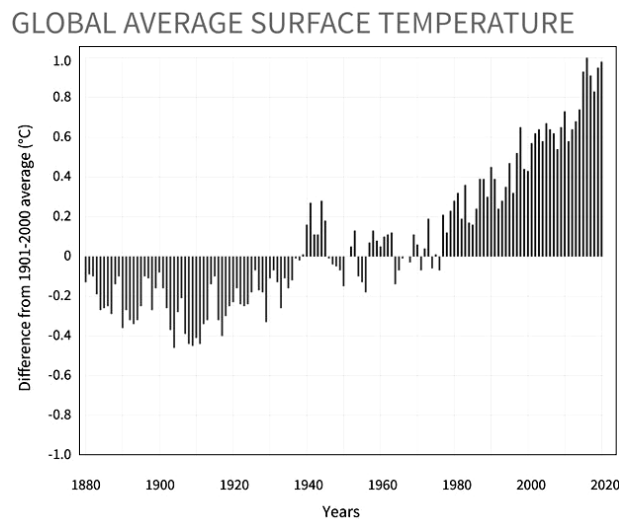


Fig. 7. Yearly surface temperature compared to the 20th-century average from 1880–2020  
Source: [36]

It is necessary to accept that influence of carbon dioxide emissions on the global temperature changing is much higher than the corresponding influence of methane. That is why the danger of global warming could be estimated by the amount of carbon dioxide emissions  $\text{CO}_2$  in metric tons per capita (CDE) [37].

In 2021, 432 natural disasters were registered, causing 10.5 thousand deaths and causing \$252 billion as economic damage [38].

For the quantitative estimation of the degree of vulnerability of the world countries to the natural disasters the index of vulnerability to natural cataclysms (NDT) is used [39]. It includes the affected from draughts, floods, hurricanes, extreme temperatures, earth-quakes and tsunami.

As a result of the control over decreasing natural resources the struggle not only between countries but also between separate groups of population can exacerbate. This process will cause new global conflicts.

Thus, the CN we calculate as follows:

$$CN = 0.3CDE + 0.7NDT.$$

### Threat 9. The state fragility (SF)

In conditions of political, social and economic instability, each country faces the task of preserving its sovereignty and improving its position in all spheres of functioning. Thus, in Fig. 8 shows the gradual growth of the World Uncertainty Index and the main reasons for its peaks [40]. It is natural that these events affect the development of each country individually.

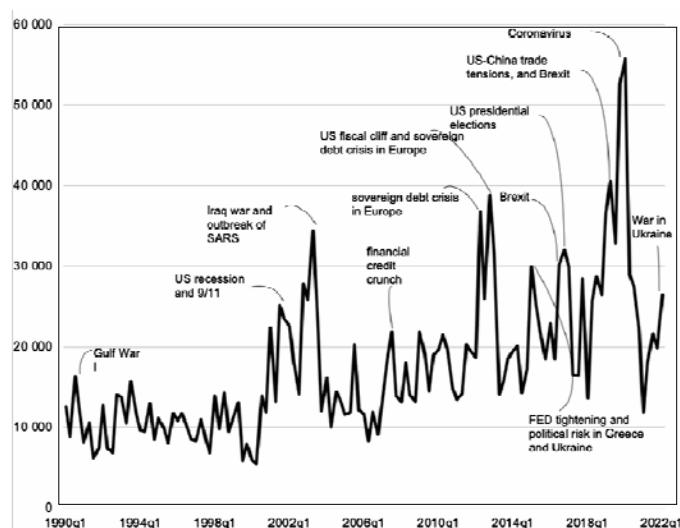


Fig. 8. World Uncertainty Index

Source: [40]

A number of such pressures act on the fragile state. For the quantitative estimation of the threat in our study the Fragile States Index produced by The Fund for Peace (FFP) is used [41]. It is based on a conflict assessment framework – known as “CAST” – that was developed by FFP nearly a quarter-century ago for assessing the vulnerability of states to collapse. The CAST framework was originally designed to measure this vulnerability and assess how it might affect projects in the field, and continues to be used widely by policy makers, field practitioners, and local community networks. The methodology uses both qualitative and quantitative indicators, relies on public source data, and produces quantifiable results.

### Threat 10. Increasing proliferation and global terrorism (PT)

This global threat we will consider in the terms of the debarment of the nuclear war, terrorism and the increasing of total number of weapons. There are three components for assessing proliferation and global terrorism:

1. The Nonproliferation Index (NPI) [42]. It defines degree of military proliferation and covers four categories of policy: demilitarization or disarmament; scientific research; state’s development; level of nonproliferation for neighbor states.

2. The Global Terrorism Index (GTI) [43]. The GTI scores each country on a scale from 0 to 10; where 0 represents no impact from terrorism and 10 represents the highest measurable impact of terrorism. It consists of:

- total number of terrorist incidents in a given year;
- total number of fatalities caused by terrorists in a given year;
- total number of injuries caused by terrorists in a given year;
- total number of hostages caused by terrorists in a given year.

3. Militarization (MLT) [44]. It is also the subindex of The Global Peace Index. It shows the state of the country armament and considers:

- military expenditure as a percentage of GDP;
- number of armed services personnel per 100.000 people;
- volume of transfers of major conventional weapons as recipient (imports) per 100.000 people;
- volume of transfers of major conventional weapons as supplier (exports) per 100.000 people;
- financial contribution to UN peacekeeping missions;
- nuclear and heavy weapons capabilities;
- ease of access to small arms and light weapons.

Thus, we obtained a comprehensive system of indicators, which is embedded in the integrated formula:

$$PT = 0.4NPI + 0.2GTI + 0.4MLT.$$

### **Threat 11. Conflict intensity increasing (CI)**

The number of military and paramilitary conflicts occurring at the national and international levels has a tendency to increase. But the nature of armed conflicts changed significantly due to the use of high-tech weapons. It should also be noted that part of the hostilities was transferred to the digital space [42].

In our research we will consider a conflict between interstate, intrastate, sub-state, and transstate ones. To assess the conflict intensity in the country we take into account the two parameters.

The first one is the levels of conflicts intensity (ICB), which were proposed Heidelberg Institute for International Conflict Research in Conflict Barometer [45]. They are dispute, non-violent crisis, violent crisis, limited war, and war. In the methodology of Conflict Barometer, the level of violence and the intensity class are considered. The last three are violent conflicts, which causes deaths and distraction of the different level.

The second is the index, which characterized the parameters of the ongoing domestic and international conflicts (OCI). It is the subindex of The Global Peace Index, which was introduced by The Institute for Economics and Peace Limited [44]. It includes:

- number and duration of internal conflicts;
- number of deaths from external/internal organized conflict;
- number, duration and role in external/internal conflicts;
- relations with neighboring countries.

The quantitative value for intensity of conflicts we will take in the form:

$$CI = 0.4ICB + 0.6OCI.$$



This threat acts as a multiplier for the security level. This naturally follows from the thesis that it is almost impossible to ensure the sustainable development of the country by participating in the armed conflicts.

Given the above, we obtained a system of indicators that identify 11 threats identified in the study. Table 1 shows their relevance to the global challenges listed in the previous section.

**Table 1.** The connection between the 11 global treats and global challenges

Global Threat		World Economic Forum	Millennium Project	United Nations
ES	The global decrease in energy security	Natural resource crises	Energy	GOAL 7: Affordable and Clean Energy
BB	The imbalance between biological capacity of the Earth and human needs in biosphere	Biodiversity loss, Human environmental damage	—	GOAL 2: Zero Hunger
GINI	Growing inequality between people and countries on the Earth	Social cohesion erosion	Population and resources, Rich-poor gap	GOAL 1: No Poverty, GOAL 10: Reduced Inequality
GD	The spread of global diseases	Infectious diseases	Health issues	GOAL 3: Good Health and Well-being
IG	Information gap	—	Global convergence of IT, Science and technology	—
CP	Corruption perception	Social cohesion erosion	Population and resources	—
WA	Limited access to drinking water	—	Clean water	GOAL 6: Clean Water and Sanitation
CN	Impact of climate change and natural disaster	Climate action failure, Extreme weather	Sustainable development and climate change	GOAL 13: Climate Action
SF	The state fragility	Livelihood crises, Debt crises	Democratization, Global foresight and decision making, Status of woman, Education and learning	GOAL 8: Decent Work and Economic Growth, GOAL 16: Peace and Justice Strong Institutions
PT	Increasing proliferation and global terrorism	—	Peace and conflict, Education and learning	GOAL 16: Peace and Justice Strong Institutions
CI	Conflict intensity increasing	Geoeconomic confrontation	Peace and conflict, Transnational organized crime, Global ethics	GOAL 16: Peace and Justice Strong Institutions

### **MODELING THE TOTAL IMPACT OF THE AGGREGATE OF 11 GLOBAL THREATS ON DIFFERENT COUNTRIES AND GROUPS OF COUNTRIES**

Let's determine the vulnerability of different countries and groups of countries to the impact of a set of 11 major threats. Quantitative data on each of the 11 threats will be obtained from the global databases specified in the description of these threats in section 2. To determine the groups of countries with close values of vulnerabilities to the impact of the 11 main threats, we use the partition algorithm

of clustering the multivariate time-series with the global alignment kernel distance [46]. It takes into the account all history of the threats and minimize differences of their values and behaviour.

Let's associate each country  $j$  with a vector  $\overline{\text{Tr}}_j$  for year  $i$ :

$$\overline{\text{Tr}}_j^i = (ES, BB, GINI, GD, IG, CP, WA, CN, SF, PT, CI)|_j^i,$$

elements of which characterize the degree of manifestation of corresponding 11 threats presented in Section 2,  $i = \overline{2005..2021}$ ,  $j = \overline{1..134}$ .

Considering the fact that all the measured data for components of vector  $\overline{\text{Tr}}_j$  are presented in different units of measurement, they have different physical meaning and vary in different ranges, they have been reduced to the normalized form, so that they vary in the range (0.1). In this case, the value 0 corresponds to the minimum value of the threat, and the value 1 corresponds to the maximum of this threat. In the study the logistic normalization is used [2].

The security index  $I_{sec}$  of each country with a value  $\|\overline{\text{Tr}}_j^i\|$  is calculated as the Minkowski norm of the vector  $\overline{\text{Tr}}_j^i$  for the  $j$ th country, composed of normalized threats. After normalization the security index  $I_{sec}$  for each country is defined as the Minkowski norm:

$$I_{sec_j}^i = \overline{S}_j^i = \left( s_{11,j}^i \sum_{k=1}^{10} (s_{k,j}^i)^p \right)^{\frac{1}{p}}$$

with parameter  $p = 3$ , where  $j = \overline{2005..2021}$ ,  $j = \overline{1..134}$ ,  $\overline{S}_j^i$  is the vector of the normilized threats  $\overline{\text{Tr}}_j^i$ .

Thus,  $I_{sec}$  defines the degree of remoteness from the influence of the set of 11 threats. Based on the calculated norms of the vector of threats  $\|\overline{\text{Tr}}_j\|$  for each country  $j$ , we obtained an order relation between clusters of countries (Table 2):

$$K_k \prec K_j \Leftrightarrow \frac{\sum_{p \in K_k} \overline{\text{Tr}}_p^{2021}}{\text{card}(K_k)} \leq \frac{\sum_{j \in K_j} \overline{\text{Tr}}_j^{2021}}{\text{card}(K_j)},$$

where  $K_k, K_j$  are the pair of obtained clusters.

From Table 2 it follows that Cluster 1 includes the group of countries most successful from the safety standpoint, for which the degree of remoteness from the set of 11 global threats is the greatest during 2005–2021. And vice versa, Cluster 5 includes the most vulnerable countries. For these countries the degree of remoteness from the set of 11 global threats is minimal.

Based on the data presented in Table 2, Fig. 9 illustrates the safety levels for different countries and regions of the world.

**Table 2.** Countries degree remoteness from the Set of Threats Based on Clustering Analysis, 2005-2021\*

CountryRank for 2021	Country, GDP per capita (constant 2015 US\$) 2020, USD**	Total influence of the set of global threats on different countries											
		(ES) The global decrease in energy security (<)	(BB) The imbalance between biological capacity of the Earth and human needs in biosphere (<)	(GINI) Growing inequality between people and countries on the Earth (>)	(GD) The spread of global diseases (<)	(IG) Information gap (<)	(CP) Corruption perception (<)	(WA) Limited access to drinking water (<)	(CN) Impact of climate change and natural disaster (<)	(SF) The state fragility (>)	(PT) Increasing proliferation and global terrorism (<)	(CI) Conflict intensity increasing (<)	(Isec) Degree of remoteness from the Set of Threats (<)
1	2	3	4	5	6	7	8	9	10	11	12	13	14

**Cluster 1 (Very high degree of remoteness during 2005–2021)**

1	Denmark (\$ 56202.17)	0.607	0.544	27.700	0.537	0.837	88.000	96.731	0.580	18.800	0.773	0.793	1.426
2	New Zealand (\$ 40218.39)	0.724	1.667	NA	0.566	0.749	88.000	100.000	0.528	18.400	0.678	0.837	1.417
4	Australia (\$ 58029.52)	3.552	1.619	34.300	0.566	0.771	73.000	NA	0.466	21.800	0.582	0.828	1.372
6	Canada (\$ 42258.69)	1.846	1.827	33.300	0.544	0.798	74.000	99.039	0.468	21.700	0.636	0.837	1.369
7	Uruguay (\$ 15044.64)	0.583	7.405	40.200	0.537	0.673	73.000	NA	0.612	35.900	0.592	0.838	1.367

**Cluster 2 (High degree of remoteness during 2005–2021)**

3	Austria (\$ 43346.43)	0.322	0.441	30.200	0.544	0.784	74.000	98.901	0.557	26.100	0.693	0.836	1.374
5	Malta (\$ 25005.76)	0.013	0.089	31.000	0.533	0.738	54.000	100.000	0.625	36.200	0.711	0.955	1.371
10	Belgium (\$ 40438.92)	0.261	0.111	27.200	0.544	0.793	73.000	99.914	0.542	31.000	0.671	0.817	1.359
14	Finland (\$ 44778.87)	0.489	1.884	27.700	0.546	0.846	88.000	99.638	0.544	70.400	0.696	0.794	1.345
15	Portugal (\$ 19771.58)	0.308	0.284	32.800	0.535	0.779	62.000	95.354	0.595	26.800	0.668	0.836	1.328
17	Czech Rep. (\$ 18984.64)	0.592	0.407	25.300	0.542	0.784	54.000	97.882	0.468	39.300	0.696	0.836	1.317
18	Slovenia (\$ 22899.36)	0.553	0.410	24.400	0.550	0.730	57.000	98.274	0.544	28.200	0.742	0.718	1.292
19	Slovakia (\$ 17360.71)	0.403	0.590	23.200	0.547	0.779	52.000	99.238	0.575	39.000	0.681	0.718	1.274
21	Germany (\$ 41315.31)	0.370	0.320	31.700	0.541	0.859	80.000	99.993	0.536	24.800	0.614	0.652	1.240
22	Italy (\$ 29359.93)	0.239	0.191	35.200	0.559	0.748	56.000	95.824	0.585	45.200	0.563	0.836	1.236
23	Latvia (\$ 15583.93)	0.557	1.273	34.500	0.555	0.760	59.000	96.289	0.612	44.000	0.656	0.717	1.231
25	Mauritius (\$ 9058.21)	0.059	0.207	36.800	0.539	0.673	54.000	NA	0.624	38.100	0.608	0.838	1.216
26	Lithuania (\$ 17213.81)	0.134	0.792	35.300	0.555	0.786	61.000	94.924	0.609	38.700	0.502	0.745	1.215
27	Estonia (\$ 19767.08)	0.105	1.165	30.800	0.538	0.836	74.000	95.761	0.498	39.500	0.640	0.657	1.207

1	2	3	4	5	6	7	8	9	10	11	12	13	14
30	Japan (\$ 34813.22)	0.137	0.129	32.900	0.537	0.786	73.000	98.565	0.480	32.200	0.603	0.658	1.183
38	Hungary (\$ 14368.69)	0.418	0.665	30.000	0.541	0.758	43.000	92.589	0.593	51.100	0.684	0.657	1.151
41	Oman (\$ 15743.22)	2.742	0.227	NA	0.587	0.619	52.000	90.557	0.479	50.400	0.484	0.718	1.125
45	South Korea (\$ 31327.41)	0.128	0.102	31.400	0.562	0.784	62.000	99.191	0.490	32.500	0.475	0.552	1.064

## Cluster 3 (Medium degree of remoteness during 2005–2021)

8	Norway (\$ 75017.16)	5.247	1.219	27.700	0.544	0.800	85.000	98.643	0.559	16.600	0.453	0.705	1.366
9	Ireland (\$ 78732.55)	0.283	0.585	30.600	0.561	0.842	74.000	97.329	0.550	22.200	0.570	0.838	1.361
11	Netherlands (\$ 46345.35)	0.317	0.137	29.200	0.564	0.871	82.000	99.972	0.534	24.100	0.407	0.836	1.358
12	Switzerland (\$ 85685.29)	0.511	0.229	33.100	0.565	0.826	84.000	94.248	0.604	19.900	0.407	0.837	1.356
13	Luxembourg (\$ 104879.26)	0.049	0.093	34.200	0.544	0.809	81.000	99.459	0.470	21.100	0.401	0.955	1.348
16	Singapore (\$ 58056.81)	0.013	0.010	NA	0.550	0.797	85.000	100.000	0.536	26.600	0.522	0.810	1.320
20	Sweden (\$ 51539.56)	0.590	1.427	29.300	0.542	0.834	85.000	99.752	0.619	21.400	0.563	0.609	1.242
24	France (\$ 35785.97)	0.539	0.538	32.400	0.563	0.806	71.000	99.249	0.597	32.500	0.307	0.701	1.227
28	Poland (\$ 14660.79)	0.630	0.396	30.200	0.536	0.755	56.000	98.325	0.540	43.100	0.574	0.746	1.204
29	Barbados (\$ 13595.03)	0.119	0.047	NA	0.555	0.494	65.000	NA	0.605	47.000	0.527	0.955	1.196
32	Spain (\$ 24939.19)	0.279	0.350	34.300	0.555	0.775	61.000	99.587	0.584	44.800	0.528	0.703	1.177
36	United Kingdom (\$ 43020.2)	0.660	0.246	35.100	0.536	0.823	78.000	99.822	0.585	41.500	0.332	0.648	1.160
46	Cyprus (\$ 26372.65)	0.046	0.053	31.200	0.544	0.762	53.000	99.765	0.575	57.400	0.528	0.606	1.064
52	Chile (\$ 12954.41)	0.256	0.777	44.900	0.555	0.640	67.000	98.771	0.494	44.100	0.414	0.613	1.021
58	United States (\$ 58203.38)	1.103	0.417	41.500	0.548	0.847	67.000	97.326	0.289	44.600	0.284	0.468	0.936

## Cluster 4 (Low degree of remoteness during 2005–2021)

31	Fiji (\$ 4911.08)	0.140	0.878	30.100	0.549	0.464	55.000	NA	0.228	16.200	0.528	0.884	1.178
33	Croatia (\$ 12984.7)	0.429	0.730	28.900	0.564	0.701	47.000	NA	0.607	49.800	0.610	0.707	1.172
34	Romania (\$ 10844.53)	0.538	0.887	34.800	0.545	0.709	45.000	81.989	0.613	51.000	0.559	0.808	1.168
35	Costa Rica (\$ 12105.93)	0.428	0.614	49.300	0.602	0.657	58.000	80.516	0.558	42.500	0.530	0.795	1.163
37	Malaysia (\$ 10631.51)	1.154	0.507	41.100	0.557	0.713	48.000	93.818	0.486	56.900	0.637	0.807	1.154
42	Cape Verde (\$ 2935.32)	0.051	0.259	42.400	0.581	0.420	58.000	NA	0.652	64.200	0.473	0.955	1.114
43	Bhutan (\$ 2879.64)	1.044	0.965	37.400	0.525	0.394	68.000	36.648	0.639	68.300	0.673	0.745	1.085
47	Argentina (\$ 11344.41)	0.916	1.866	42.300	0.568	0.570	38.000	NA	0.588	50.100	0.482	0.767	1.062
48	Guyana (\$ 9250.3)	0.005	21.338	NA	0.555	0.430	39.000	NA	0.490	66.100	0.642	0.606	1.061
49	Botswana (\$ 6299.21)	0.589	1.394	53.300	0.421	0.560	55.000	NA	0.609	57.000	0.570	0.838	1.058

1	2	3	4	5	6	7	8	9	10	11	12	13	14
51	Kazakhstan (\$ 10974.96)	2.603	0.722	27.800	0.568	0.607	37.000	89.335	0.496	61.200	0.495	0.606	1,029
53	Serbia (\$ 6549.35)	0.877	0.547	34.500	0.558	0.671	38.000	75.038	0.549	67.400	0.630	0.606	0.989
54	Moldova (\$ 3250.31)	0.023	0.667	26.000	0.540	0.583	36.000	74.071	0.611	67.000	0.598	0.552	0.964
55	North Macedonia (\$ 5115.92)	0.395	0.482	33.000	0.480	0.622	39.000	76.833	0.619	64.500	0.645	0.564	0.950
56	Montenegro (\$ 6512.62)	0.665	0.594	36.800	0.562	0.592	46.000	85.072	0.608	58.500	0.515	0.564	0.949
57	Dominican Republic (\$ 7677.71)	0.060	0.378	39.600	0.563	0.567	30.000	NA	0.448	64.700	0.603	0.718	0.942
59	Viet Nam (\$ 2655.77)	0.734	0.413	35.700	0.537	0.677	39.000	NA	0.389	63.300	0.505	0.658	0.931
60	Jamaica (\$ 4539)	0.039	0.256	45.500	0.561	0.498	44.000	NA	0.570	61.200	0.625	0.613	0.920
61	Ghana (\$ 2018.62)	1.479	0.648	43.500	0.451	0.588	43.000	41.410	0.596	63.900	0.577	0.656	0.919
62	Jordan (\$ 4028.96)	0.055	0.086	33.700	0.539	0.487	49.000	85.701	0.637	76.800	0.582	0.543	0.911
64	Greece (\$ 17323.82)	0.280	0.285	33.100	0.537	0.700	49.000	100.000	0.567	54.500	0.459	0.459	0.896
65	Albania (\$ 4389.9)	0.395	0.526	30.800	0.564	0.549	35.000	70.675	0.363	59.000	0.549	0.564	0.887
66	Benin (\$ 1214.66)	0.000	0.572	37.800	0.431	0.476	42.000	NA	0.637	72.800	0.501	0.690	0.886
67	Cote D'Ivoire (\$ 2313.79)	0.877	1.281	37.200	0.433	0.523	36.000	35.205	0.669	90.700	0.602	0.596	0.885
68	Bosnia and Herzegovina (\$ 5433.15)	0.609	0.514	33.000	0.531	0.499	35.000	88.869	0.464	72.900	0.548	0.611	0.885
71	Senegal (\$ 1364.84)	0.048	0.785	38.100	0.531	0.432	43.000	NA	0.594	73.400	0.626	0.561	0.863
72	Tanzania (\$ 1061.17)	0.415	0.839	40.500	0.420	0.501	39.000	NA	0.533	79.300	0.604	0.698	0.860
74	Paraguay (\$ 5670.75)	1.007	3.278	43.500	0.593	0.516	30.000	64.084	0.323	66.400	0.449	0.564	0.855
75	Indonesia (\$ 3756.91)	1.971	0.722	37.300	0.519	0.595	38.000	NA	0.602	67.600	0.480	0.505	0.854
76	Tunisia (\$ 3780.6)	0.303	0.339	32.800	0.537	0.626	44.000	79.286	0.636	69.200	0.497	0.455	0.849
77	Armenia (\$ 4021.05)	0.291	0.324	25.200	0.599	0.537	49.000	86.911	0.611	69.800	0.454	0.388	0.844
78	Azerbaijan (\$ 5083.38)	4.071	0.382	26.600	0.565	0.575	30.000	88.323	0.621	75.100	0.460	0.352	0.839
79	Ecuador (\$ 5317.68)	1.862	1.108	47.300	0.576	0.427	36.000	66.827	0.536	71.200	0.525	0.560	0.839
80	Guinea (\$ 984.01)	0.136	1.116	29.600	0.447	0.348	25.000	NA	0.667	97.400	0.506	0.558	0.838
81	Georgia (\$ 4447.66)	0.300	0.545	34.500	0.530	0.641	55.000	66.355	0.611	72.600	0.476	0.459	0.835
82	Sierra Leone (\$ 623.89)	0.102	0.880	35.700	0.462	0.361	34.000	10.621	0.661	83.400	0.631	0.562	0.835
84	Kyrgyzstan (\$ 1100.04)	0.659	0.779	29.000	0.582	0.488	27.000	70.090	0.644	76.400	0.748	0.365	0.832
85	Belarus (\$ 6234.82)	0.081	0.649	24.400	0.556	0.565	41.000	94.611	0.561	68.000	0.526	0.362	0.805
86	Algeria (\$ 3834.44)	2.702	0.243	27.600	0.571	0.415	33.000	72.381	0.603	73.600	0.413	0.389	0.794

1	2	3	4	5	6	7	8	9	10	11	12	13	14
88	Belize (\$ 3968.49)	0.536	0.446	53.300	0.571	0.447	NA	NA	0.609	64.200	0.436	0.500	0.789
89	Brazil (\$ 8228.78)	0.945	3.327	48.900	0.551	0.627	38.000	85.766	0.545	75.800	0.605	0.367	0.783
90	El Salvador (\$ 3632.45)	0.219	0.290	38.800	0.587	0.434	34.000	NA	0.425	71.600	0.464	0.563	0.769
91	Morocco (\$ 2818.77)	0.069	0.450	39.500	0.561	0.589	39.000	79.950	0.530	71.500	0.511	0.443	0.766
94	Gambia (\$ 692.21)	0.003	0.496	35.900	0.518	0.400	37.000	44.715	0.562	80.500	0.507	0.563	0.752
97	Ukraine (\$ 2344.36)	0.774	1.130	25.600	0.523	0.622	32.000	89.020	0.607	69.800	0.447	0.273	0.716
98	Egypt (\$ 4028.42)	0.898	0.199	31.500	0.576	0.558	33.000	NA	0.637	85.000	0.400	0.322	0.712
101	Tajikistan (\$ 1199.06)	0.844	0.422	34.000	0.589	0.323	25.000	55.237	0.652	75.100	0.628	0.302	0.705
105	Colombia (\$ 5892.48)	2.825	1.852	54.200	0.574	0.548	39.000	73.009	0.601	79.300	0.269	0.291	0.685
108	Israel (\$ 37488.45)	0.379	0.039	38.600	0.539	0.770	59.000	99.321	0.550	75.100	0.249	0.237	0.676
109	Nicaragua (\$ 1922.35)	0.244	1.537	46.200	0.575	0.336	20.000	55.516	0.404	77.100	0.625	0.404	0.675
111	Russian (\$ 9666.81)	2.449	1.265	35.300	0.529	0.689	29.000	76.104	0.503	73.600	0.217	0.264	0.655
113	Iran (\$ 4883.6)	1.604	0.223	40.900	0.582	0.589	25.000	93.984	0.244	84.500	0.318	0.285	0.632
114	Saudi Arabia (\$ 18691.25)	3.415	0.082	NA	0.606	0.620	53.000	NA	0.478	69.700	0.369	0.167	0.630
116	Venezuela (\$ NA)	2.310	1.154	44.800	0.576	0.373	14.000	NA	0.594	92.600	0.386	0.292	0.619
117	Mexico (\$ 8909.68)	0.683	0.485	45.400	0.644	0.612	31.000	43.026	0.572	69.900	0.450	0.197	0.601
118	Lebanon (\$ 5382.34)	0.010	0.085	31.800	0.564	0.413	24.000	47.700	0.496	89.000	0.262	0.299	0.565
121	Turkey (\$ 12038.63)	0.306	0.390	41.900	0.566	0.616	38.000	NA	0.591	79.700	0.404	0.145	0.530
129	Burkina Faso (\$ 731.52)	0.034	0.792	47.300	0.473	0.416	42.000	NA	0.535	87.100	0.542	0.185	0.472
130	Cameroon (\$ 1419.68)	1.585	1.313	46.600	0.416	0.393	27.000	NA	0.659	97.200	0.448	0.113	0.464
132	Syria (\$ NA)	0.422	0.364	37.500	0.592	0.344	13.000	NA	0.607	110.700	0.325	0.059	0.440
133	Yemen (\$ NA)	0.803	0.737	36.700	0.579	0.234	16.000	NA	0.595	111.700	0.280	0.066	0.438

## Cluster 5 (Very low degree of remoteness during 2005–2021)

39	Mongolia (\$ 4126.7)	5.942	1.944	32.700	0.539	0.461	35.000	30.061	0.208	52.300	0.574	0.795	1.137
50	Congo (\$ 1608.78)	9.043	8.538	48.900	0.411	0.303	21.000	45.897	0.432	92.400	0.479	0.661	1.045
63	Zambia (\$ 1273.88)	0.681	1.373	57.100	0.403	0.455	33.000	NA	0.645	84.900	0.621	0.694	0.905
69	Angola (\$ 3168.25)	9.381	2.131	51.300	0.403	0.307	29.000	NA	0.534	89.000	0.563	0.506	0.873
70	Bolivia (\$ 2983.03)	2.081	4.583	43.600	0.560	0.392	30.000	NA	0.375	74.900	0.515	0.564	0.873
73	Namibia (\$ 4047.86)	0.134	2.519	59.100	0.407	0.395	49.000	NA	0.307	64.300	0.531	0.795	0.859
83	Madagascar (\$ 442.19)	0.160	2.324	42.600	0.492	0.328	26.000	20.539	0.389	79.500	0.643	0.739	0.832

1	2	3	4	5	6	7	8	9	10	11	12	13	14
87	Peru (\$ 5792.19)	0.980	1.525	43.800	0.578	0.529	36.000	51.264	0.471	71.400	0.402	0.560	0.792
92	Thailand (\$ 6198.41)	0.456	0.540	35.000	0.566	0.668	35.000	NA	0.303	70.900	0.501	0.430	0.763
93	Nepal (\$ 1028.46)	0.252	0.519	32.800	0.559	0.453	33.000	17.576	0.416	82.200	0.499	0.549	0.754
95	China (\$ 10370.36)	0.812	0.243	38.200	0.558	0.684	45.000	NA	0.273	68.900	0.390	0.427	0.745
96	Laos (\$ 2554.43)	1.191	0.922	38.800	0.573	0.297	30.000	17.682	0.258	76.000	0.475	0.564	0.732
99	Togo (\$ 626.62)	0.002	0.569	42.400	0.447	0.369	30.000	19.561	0.661	85.100	0.418	0.549	0.707
100	Malawi (\$ 394)	0.455	0.767	38.500	0.439	0.353	35.000	NA	0.345	83.200	0.501	0.694	0.707
102	Rwanda (\$ 834.39)	0.168	0.516	43.700	0.413	0.449	53.000	12.103	0.663	85.000	0.452	0.445	0.701
103	Guatemala (\$ 4126.21)	0.263	0.530	48.300	0.598	0.371	25.000	55.834	0.283	79.400	0.471	0.511	0.691
104	Sri Lanka (\$ 4052.75)	0.157	0.300	39.300	0.597	0.487	37.000	NA	0.267	80.500	0.421	0.446	0.686
106	Honduras (\$ 2223.45)	0.283	1.103	48.200	0.593	0.290	23.000	NA	0.404	79.400	0.456	0.503	0.682
107	South Africa (\$ 5659.21)	1.025	0.262	63.000	0.355	0.642	44.000	NA	0.451	70.000	0.534	0.416	0.682
110	Lesotho (\$ 982.97)	0.279	0.516	44.900	0.375	0.339	38.000	28.906	0.243	77.900	0.567	0.641	0.673
112	Bangladesh (\$ 1643.67)	0.694	0.459	32.400	0.513	0.492	26.000	58.512	0.315	85.000	0.515	0.388	0.636
115	Cambodia (\$ 1376.41)	0.204	0.796	NA	0.491	0.354	23.000	27.758	0.375	80.600	0.488	0.604	0.626
119	Uganda (\$ 891.3)	0.343	0.441	42.700	0.378	0.422	27.000	16.648	0.639	92.900	0.513	0.316	0.564
120	India (\$ 1811.68)	0.523	0.368	35.700	0.487	0.590	40.000	NA	0.358	77.000	0.382	0.250	0.534
122	Mozambique (\$ 574.6)	3.192	1.962	54.000	0.292	0.312	26.000	NA	0.315	93.900	0.481	0.256	0.520
123	Nigeria (\$ 2396.04)	3.467	0.632	35.100	0.368	0.497	24.000	21.669	0.619	98.000	0.543	0.108	0.517
124	Pakistan (\$ 1446.81)	0.507	0.433	29.600	0.479	0.424	28.000	35.839	0.508	90.500	0.239	0.233	0.504
125	Kenya (\$ 1559.55)	0.238	0.468	40.800	0.434	0.514	30.000	NA	0.335	89.200	0.391	0.342	0.492
126	CAR (\$ 414.4)	0.174	6.237	56.200	0.369	0.239	24.000	6.183	0.648	107.000	0.467	0.094	0.491
127	Philippines (\$ 3269.67)	0.315	0.351	42.300	0.476	0.628	33.000	47.465	0.205	82.400	0.390	0.230	0.482
128	Niger (\$ 522.56)	0.785	0.843	37.300	0.482	0.302	31.000	NA	0.343	96.000	0.374	0.318	0.482
31	Mali (\$ 781.73)	0.222	1.151	36.100	0.444	0.353	29.000	NA	0.616	96.600	0.408	0.126	0.448
134	Ethiopia (\$ 826.95)	0.376	0.547	35.000	0.474	0.332	39.000	12.577	0.523	99.000	0.417	0.111	0.403

NA – data not available; critical threats are indicated by red color;

(<) – a lower value corresponds to a higher threat. (>) – a higher value corresponds to a higher threat;

\*latest available data;

\*\* Data Source: [47].

As presented in Table 3, the common trait of the ten leaders is high *Isec*, and low level of threats. E.g., the group leaders, Denmark, New Zealand, and Austria, have the best indicators among all the group countries. However, half of the list have a relatively low level of energy security (ES), this threat is critical for them. It is also necessary to pay attention to the low biological balance (BB) of Belgium and Malta, and the high level of inequality (GINI) of Uruguay.



Fig. 9. Countries safety levels – degree of remoteness from the Set of Threats (Based on Clustering Analysis)

Table 3. Top 10 countries with the highest level of national security\*

Rank Isec	ISO	Country	Isec	ES	BB	GINI	GD	IG	CP	WA	CN	SF	PT	CI
1	DNK	Denmark	1.426	0.607	0.544	27.700	0.537	0.837	88.000	96.731	0.580	18.800	0.773	0.793
2	NZL	New Zealand	1.417	0.724	1.667	NA	0.566	0.749	88.000	100.000	0.528	18.400	0.678	0.837
3	AUT	Austria	1.374	0.322	0.441	30.200	0.544	0.784	74.000	98.901	0.557	26.100	0.693	0.836
4	AUS	Australia	1.372	3.552	1.619	34.300	0.566	0.771	73.000	NA	0.466	21.800	0.582	0.828
5	MLT	Malta	1.371	0.013	0.089	31.000	0.533	0.738	54.000	100.000	0.625	36.200	0.711	0.955
6	CAN	Canada	1.369	1.846	1.827	33.300	0.544	0.798	74.000	99.039	0.468	21.700	0.636	0.837
7	URY	Uruguay	1.367	0.583	7.405	40.200	0.537	0.673	73.000	NA	0.612	35.900	0.592	0.838
8	NOR	Norway	1.366	5.247	1.219	27.700	0.544	0.800	85.000	98.643	0.559	16.600	0.453	0.705
9	IRL	Ireland	1.361	0.283	0.585	30.600	0.561	0.842	74.000	97.329	0.550	22.200	0.570	0.838
10	BEL	Belgium	1.359	0.261	0.111	27.200	0.544	0.793	73.000	99.914	0.542	31.000	0.671	0.817

\* For each country critical values of the threats indicators are highlighted by red color

The G7 countries are characterized by a high and medium level of national security and therefore a low vulnerability to the impact of 11 global threats (Table 4). In some sense, an exception is the United States, for which the threats of inequality (GINI), global warming and natural disasters (CN), conflicts (CI), and the proliferation (PT) are very critical. In Japan, there are clearly threats to disrupt the biological balance (BB) and energy security (ES), which is natural in connection with the geographic location and the great density of the population.

Table 4. The level of national security of the G-7 countries\*

Rank Isec	ISO	Country	Isec	ES	BB	GINI	GD	IG	CP	WA	CN	SF	PT	CI
6	CAN	Canada	1.369	1.846	1.827	33.300	0.544	0.798	74.000	99.039	0.468	21.700	0.636	0.837
21	DEU	Germany	1.240	0.370	0.320	31.700	0.541	0.859	80.000	99.993	0.536	24.800	0.614	0.652
22	ITA	Italy	1.236	0.239	0.191	35.200	0.559	0.748	56.000	95.824	0.585	45.200	0.563	0.836
24	FRA	France	1.227	0.539	0.538	32.400	0.563	0.806	71.000	99.249	0.597	32.500	0.307	0.701
30	JPN	Japan	1.183	0.137	0.129	32.900	0.537	0.786	73.000	98.565	0.480	32.200	0.603	0.658
36	GBR	United Kingdom	1.160	0.660	0.246	35.100	0.536	0.823	78.000	99.822	0.585	41.500	0.332	0.648
58	USA	United States	0.936	1.103	0.417	41.500	0.548	0.847	67.000	97.326	0.289	44.600	0.284	0.468

\* For each country critical values of the threats indicators are highlighted by red color



The last 10 countries by the *Isec* are characterized by a very high vulnerability to the impact of 11 global threats (Table 5). These are countries with hostilities on their territory (CI), low stability of the state (SF) and a high level of proliferation (PT).

**Table 5.** Last 10 countries with the lowest level of national security\*

Rank Isec	ISO	Country	Isec	ES	BB	GINI	GD	IG	CP	WA	CN	SF	PT	CI
125	KEN	Kenya	0,492	0,238	0,468	40,800	0,434	0,514	30,000	NA	<b>0,335</b>	89,200	0,391	0,342
126	CAF	Central African Republic	0,491	0,174	6,237	<b>56,200</b>	<b>0,369</b>	<b>0,239</b>	24,000	<b>6,183</b>	0,648	<b>107,000</b>	0,467	<b>0,094</b>
127	PHL	Philippines	0,482	0,315	0,351	42,300	0,476	0,628	33,000	47,465	<b>0,205</b>	82,400	0,390	<b>0,230</b>
128	NER	Niger	0,482	0,785	0,843	37,300	0,482	<b>0,302</b>	31,000	NA	<b>0,343</b>	<b>96,000</b>	0,374	<b>0,318</b>
129	BFA	Burkina Faso	0,472	0,034	0,792	<b>47,300</b>	0,473	0,416	42,000	NA	0,535	<b>87,100</b>	0,542	<b>0,185</b>
130	CMR	Cameroon	0,464	1,585	1,313	<b>46,600</b>	<b>0,416</b>	0,393	27,000	NA	0,659	<b>97,200</b>	0,448	<b>0,113</b>
131	MLI	Mali	0,448	0,222	1,151	36,100	0,444	0,353	29,000	NA	0,616	<b>96,600</b>	0,408	<b>0,126</b>
132	SYR	Syria	0,440	0,422	0,364	37,500	0,592	<b>0,344</b>	<b>13,000</b>	NA	0,607	<b>110,700</b>	<b>0,325</b>	<b>0,059</b>
133	YEM	Yemen	0,438	0,803	0,737	36,700	0,579	<b>0,234</b>	<b>16,000</b>	NA	0,595	<b>111,700</b>	<b>0,280</b>	<b>0,066</b>
134	ETH	Ethiopia	0,403	0,376	0,547	35,000	0,474	0,332	39,000	<b>12,577</b>	0,523	<b>99,000</b>	0,417	<b>0,111</b>

\* For each country critical values of the threats indicators are highlighted by red color

As for Ukraine, the most critical for it is the threat of increasing armed conflicts (CI), which prevents the sustainable development of its territory (Table 6). Among the neighboring countries, Ukraine ranks second to last (the last is Russia).

**Table 6.** Ukraine in the European context\*

Rank Isec	ISO	Country	Isec	ES	BB	GINI	GD	IG	CP	WA	CN	SF	PT	CI
19	SVK	Slovakia	1,274	0,403	0,590	23,200	0,547	0,779	52,000	99,238	0,575	39,000	0,681	0,718
21	DEU	Germany	1,240	0,370	<b>0,320</b>	31,700	0,541	0,859	80,000	99,993	0,536	24,800	0,614	0,652
22	ITA	Italy	1,236	<b>0,239</b>	<b>0,191</b>	<b>35,200</b>	0,559	0,748	56,000	95,824	0,585	45,200	0,563	0,836
24	FRA	France	1,227	0,539	0,538	32,400	0,563	0,806	71,000	99,249	0,597	32,500	<b>0,307</b>	0,701
28	POL	Poland	1,204	0,630	0,396	30,200	<b>0,536</b>	0,755	56,000	98,325	0,540	43,100	0,574	0,746
34	ROU	Romania	1,168	0,538	0,887	34,800	0,545	0,709	45,000	81,989	0,613	51,000	0,559	0,808
36	GBR	United Kingdom	1,160	0,660	0,246	35,100	0,536	0,823	78,000	99,822	0,585	41,500	<b>0,332</b>	0,648
38	HUN	Hungary	1,151	0,418	0,665	30,000	0,541	0,758	<b>43,000</b>	92,589	0,593	51,100	0,684	0,657
54	MDA	Moldova, Republic of	0,964	<b>0,023</b>	0,667	26,000	0,540	0,583	36,000	74,071	0,611	67,000	0,598	0,552
85	BLR	Belarus	0,805	<b>0,081</b>	0,649	24,400	0,556	0,565	41,000	94,611	0,561	68,000	0,526	<b>0,362</b>
97	UKR	Ukraine	0,716	0,774	1,130	25,600	0,523	0,622	32,000	89,020	0,607	69,800	0,447	<b>0,273</b>
111	RUS	Russian Federation	0,655	2,449	1,265	35,300	0,529	0,689	29,000	76,104	0,503	73,600	<b>0,217</b>	<b>0,264</b>

\* For each country critical values of the threats indicators are highlighted by red color

### POSSIBLE SCENARIOS OF WORLD DEVELOPMENT DURING “CONFLICT XXI”

According to the results of the above-mentioned studies, the following scenarios of the development of global society during the conflict of the 21st century and after its end can be assumed:

**Pessimistic scenario.** As a result of the conducted research, the question arises: what does the 21st century have in store for civilization? What is the nature of the final state of civilization as a system? What should happen to world civilization, in particular, in the 22nd century? Perhaps the final cycle of some global evolutionary chain of human development is beginning?

The answer to this question can be found in the research of two outstanding scientists of the last century: Vernadskyi [48] and Moiseev [49]. Independently of each other, they formulated a very close idea: if humanity on a planetary scale does not radically change its behaviour (using its mind and its work to self-destruct), then in the middle of the 21st century conditions may arise in which people will not be able to exist. Such conclusions were made for the paradigm constant throughout the history of mankind: “unlimited and growing consumption” and for the technosphere (a set of technological ways of life), unfriendly to human habitation, which developed in the 19th and early 21st centuries.

**An optimistic scenario.** If humanity can change the paradigm of its behaviour on a planetary scale, for example, to “harmonious coexistence” and radically transform the technosphere into a “nature-like” one (a human-friendly environment based on the convergence of nano-, bio-informational, cognitive and socio-humanitarian technologies [50]), then the regularity revealed for the previous paradigm of the development of systemic world conflicts is not justified for the new paradigm. And this, in turn, will allow humanity to continue its creative mission on planet Earth.

## CONCLUSIONS

The study analyzed the global challenges and problems of humanity. According to the results of the analysis, the set of global threats to sustainable development have been defined.

Based on the provided global modelling, 5 clusters of countries were identified according to their remoteness from these threats during 2005–2021. Using the obtained results, critical threats were identified for each country. Groups of countries (Top10, Last10, G7, Ukraine in the European context) were analyzed and their features were identified.

Assumptions are made about possible pessimistic and optimistic scenarios for the development of the world during the “Conflict XXI” and after it.

These results can be used to study and model the degree of cultural and civilizational gaps in the world.

## REFERENCES

1. M. Zgurovsky, A. Boldak, K. Yefremov, and I. Pyshnograiev, “Modeling and investigating the behavior of complex socio-economic systems,” *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, pp. 1113–1116. Available: <https://doi.org/10.1109/UKRCON.2017.8100400>.
2. M.Z. Zgurovsky and Y.P. Zaychenko, “Intellectual Analysis of Systemic World Conflicts and Global Forecast for the 21st Century,” in *Big Data: Conceptual Analysis and Applications. Studies in Big Data*, vol. 58, Springer, Cham, 2020, pp. 141–277. Available: [https://doi.org/10.1007/978-3-030-14298-8\\_4](https://doi.org/10.1007/978-3-030-14298-8_4).

3. *The Global Risks Report 2022*, 17th Edition, WEF, 2022. Available: [https://www3.weforum.org/docs/WEF\\_The\\_Global\\_Risks\\_Report\\_2022.pdf](https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2022.pdf).
4. "The 15 Global Challenges," *The Millennium Project*. Accessed on: May 23, 2022. [Online]. Available: <https://www.millennium-project.org/projects/challenges/>.
5. *Transforming our world: The 2030 agenda for sustainable development*. United Nations, New York, NY, 2015. Accessed on: May 23, 2022. [Online]. Available: <https://sdgs.un.org/2030agenda>.
6. "Sustainable Development Goals kick off with start of new year," *UN*. Accessed on: May 23, 2022. [Online]. Available: <https://www.un.org/sustainabledevelopment/log/2015/12/sustainable-development-goals-kick-off-with-start-of-new-year/>.
7. "Global Energy Transition Statistics", *Enerdata*. Accessed on: May 23, 2022. [Online]. Available: <https://yearbook.enerdata.net>.
8. "Short-term energy outlook?", *EIA*. Release Date: May 10, 2022. [Online]. Available: <https://www.eia.gov/outlooks/steo/>.
9. "Energy Outlook 2022", *BP*. [Online]. Available: <https://www.bp.com/content/am/bp/usinss-sites/en/global/corporate/pdfs/energy-economics/energy-outlook/bp-energy-outlook-2022.pdf>.
10. M. Zgurovsky, Ed., *Foresight: challenges to the energy independence of countries and regions of the world in the medium-term (until 2025) and long-term (until 2030) time horizons*. Kyiv: Igor Sikorsky Kyiv Polytechnic Institute, "Polytechnica", 2021, 188 p. [Online]. Available: <http://wdc.org.ua/sites/default/files/Foresight-Energy-ua-2021.pdf>.
11. "World population," *Worldometer*. Accessed on: May 23, 2022. [Online]. Available: <https://www.worldometers.info>.
12. "Country Trends," *Global Footprint Network*. Accessed on: May 23, 2022. [Online]. Available: [https://data.footprintnetwork.org/?\\_ga=2.48112801.784415924.1653213413-1391004579.1653213413#/countryTrends?cn=5001&type=Ctot,EFctot](https://data.footprintnetwork.org/?_ga=2.48112801.784415924.1653213413-1391004579.1653213413#/countryTrends?cn=5001&type=Ctot,EFctot).
13. "Ecological Footprint," *Global Footprint Network*. Accessed on: May 23, 2022. [Online]. Available: <https://www.footprintnetwork.org/our-work/ecological-footprint/>.
14. "Poverty and Inequality Platform," *The World Bank*. Accessed on: May 23, 2022. [Online]. Available: <https://pip.worldbank.org/home>.
15. R. Andres et al., "April 2022 global poverty update from the World Bank," *World Bank Blogs*. [Online]. Available: <https://blogs.worldbank.org/opendata/april-2022-global-poverty-update-world-bank>.
16. N. Yonzan, C. Lakner, and D.G. Mahler, "Is COVID-19 increasing global inequality?" *World Bank Blogs*. [Online]. Available: <https://blogs.worldbank.org/opendata/covid-19-increasing-global-inequality>.
17. "GINI index," *The World Bank Data*. Accessed on: May 23, 2022. [Online]. Available: <https://data.worldbank.org/indicator/SI.POV.GINI>.
18. "The top 10 causes of death," *WHO*. Accessed on: May 23, 2022. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
19. "Noncommunicable diseases," *WHO*. Accessed on: May 23, 2022. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>.
20. "Malaria," *WHO*. Accessed on: May 23, 2022. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/malaria>.
21. "HIV/AIDS," *WHO*. Accessed on: May 23, 2022. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/hiv-aids>.
22. "Foresight COVID-19: impact on economy and society," *World Data Center "Geoinformatics and Sustainable Development"*. Accessed on: May 23, 2022. [Online]. Available: <http://wdc.org.ua/en/node/190017>.
23. "The global health observatory," *WHO*. Accessed on: May 23, 2022. [Online]. Available: <https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates>.

24. O. Djuraskovic, "Big Data Statistics 2022: How Much Data is in The World?" *FSG*. Accessed on: May 23, 2022. [Online]. Available: <https://firstsiteguide.com/big-data-stats/>.
25. "The ICT Development Index (IDI): conceptual framework and methodology," *ITU*. Accessed on: May 23, 2022. [Online]. Available: <https://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2016/methodology.aspx>.
26. "Statistics," *ITU*. Accessed on: May 23, 2022. [Online]. Available: <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>.
27. "ICT goods exports (% of total goods exports)," *The World Bank Data*. Accessed on: May 23, 2022. [Online]. Available: <https://data.worldbank.org/indicator/X.VAL.ICTG.ZS.UN>.
28. "ICT service exports (BoP, current US\$)," *The World Bank Data*. Accessed on: May 23, 2022. [Online]. Available: <https://data.worldbank.org/indicator/BX.GSR.FSV.CD>
29. "Internet servers (per 1 million people)," *The World Bank Data*. Accessed on: May 23, 2022. [Online]. Available: <https://data.worldbank.org/indicator/IT.ET.SECR.P6>.
30. "Global Cybersecurity Index," *ITU*. Accessed on: May 23, 2022. [Online]. Available: <https://www.itu.int/epublications/publication/D-STR-GCI.01-2021-htm-e/>.
31. "World Press Freedom index," *Reporters Without Borders*. Accessed on: May 23, 2022. [Online]. Available: <https://rsf.org/en/index>.
32. "CPI 2021: HIGHLIGHTS AND INSIGHTS," *Transparency International*. Accessed on: May 23, 2022. [Online]. Available: <https://www.transparency.org/en/ews/pi-2021-highlights-insights>.
33. "CORRUPTION PERCEPTIONS INDEX," *Transparency International*. Accessed on: May 23, 2022. [Online]. Available: <https://www.transparency.org/cpi>.
34. "Progress on household drinking water, sanitation and hygiene, 2000-2020", UNICEF. [Online]. Available: <https://data.unicef.org/resources/progress-on-household-drinking-water-sanitation-and-hygiene-2000-2020/> [Accessed May 23, 2022].
35. "People using safely managed drinking water services (% of population)," *WHO/UNICEF Joint Monitoring Programme, The World Bank Data*. Accessed on: May 23, 2022. [Online]. Available: <https://data.worldbank.org/indicator/H.H2O.SMDW.ZS>.
36. R. Lindsey and L. Dahlman, "Climate Change: Global Temperature," NOAA *Climate.gov*. Accessed on: May 23, 2022. [Online]. Available: <https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature>.
37. "CO2 emissions (metric tons per capita)," *The World Bank Data*. Accessed on: May 23, 2022. [Online]. Available: <https://data.worldbank.org/indicator/EN.ATM.CO2E.PC>
38. "2021 Disasters in numbers," *CRED*, 2022. [Online]. Available: [https://cred.be/sites/default/files/2021\\_EMDAT\\_report.pdf](https://cred.be/sites/default/files/2021_EMDAT_report.pdf).
39. M. Zgurovsky, Ed., *Foresight 2018: systemic world conflicts and global forecast for XXI*, K.: NTUU "Igor Sikorsky Kyiv Polytechnic Institute", 2018, 226 p. [Online]. Available: <http://wdc.org.ua/sites/default/files/WDC-IASA-FORESIGHT-2018-EN.pdf>.
40. H. Ahir, N. Bloom, and D. Furceri, "World Uncertainty Index," *Stanford mimeo*, 2018. Accessed on: May 23, 2022. [Online]. Available: <https://worlduncertaintyindex.com>.
41. "Fragile State Index," *The Fund for Peace*. Accessed on: May 23, 2022. [Online]. Available: <https://fragilestatesindex.org>.
42. M. Zgurovsky, M. Kravchenko, I. Pyshnograiev, and M. Perestyuk, "Modeling of the intercivilization fault effect on the conflict intensity throughout the world," *System Research & Information Technologies*, no. 4, 2021. [Online]. Available: <http://journal.iasa.kpi.ua/article/view/251997>.
43. "Global Terrorism Index," *Institute for Economics and Peace*. Accessed on: May 23, 2022. [Online]. Available: <https://www.visionofhumanity.org/maps/global-terrorism-index/#/>.

44. "Global Peace Index," *Institute for Economics and Peace*. Accessed on: May 23, 2022. [Online]. Available: <https://www.visionofhumanity.org/maps/#/>.
45. "Conflict Barometer," *HIK*. Accessed on: May 23, 2022. [Online]. Available: <https://hiik.de/conflict-barometer/current-version/?lang=en>.
46. A. Sardá-Espinosa, "Time-Series Clustering in R Using the dtwclust Package," *The R Journal*, vol. 11/01, 2019. [Online]. Available: <https://journal.r-project.org/archive/2019/RJ-2019-023/RJ-2019-023.pdf>.
47. "GDP per capita (constant 2015 US\$)," *The World Bank*. Accessed on: May 23, 2022. [Online]. Available: <https://data.worldbank.org/indicator/NY.GDP.PCAP.KD>
48. V. Vernadskii, "A few words on the noosphere," *Uspekhi Sovrem. Biologii*, 18 (2), 1944.
49. N. Moiseyev, "Save mankind on the Earth," *Ekologiya i Zhizn*, no. 1, pp. 11–13, 2000.
50. M. Kovalchuk, "Science and Life: My Convergence," *Autobiographical Sketches: Science Educational and Conceptual Articles*, vol. 1, 2011.

*Received 22.07.2022*

#### INFORMATION ON THE ARTICLE

**Ivan O. Pyshnograiev**, ORCID: 0000-0002-3346-8318, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: [pyshnograiev@gmail.com](mailto:pyshnograiev@gmail.com)

**Michael Z. Zgurovsky**, ORCID: 0000-0001-5896-7466, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: [zgurovsm@hotmail.com](mailto:zgurovsm@hotmail.com)

#### ДОСЛІДЖЕННЯ ТЕНДЕНЦІЙ БЕЗПЕКИ ГЛОБАЛЬНОГО СУСПІЛЬСТВА НА ОСНОВІ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ / М.З. Згуровський, І.О. Пишнограєв

**Анотація.** Присвячено застосуванню методології системного аналізу та інтелектуального аналізу даних до однієї з найактуальніших проблем сучасності: дослідження безпеки глобального суспільства в конфліктному світі. Розглянуто комплекс глобальних загроз, актуальних для першої половини ХХІ ст. Ці загрози були визначені Організацією Об'єднаних Націй, Всесвітньою організацією охорони здоров'я, Всесвітнім економічним форумом та іншими авторитетними міжнародними організаціями. У результаті застосування методу Delphi для аналізу широкого спектру загроз, виявлених цими організаціями, виявлено 11 найважливіших загроз людству в першій половині ХХІ ст. Проаналізовано вразливість різних країн до впливу сукупності цих загроз. Побудовано сценарії можливого розвитку глобального суспільства під час та після конфлікту.

**Ключові слова:** глобальна безпека, системні конфлікти, глобальні загрози, норма Мінковського, вразливість.

## **SUPERCONDUCTING GRAVIMETERS BASED ON ADVANCED NANOMATERIALS AND QUANTUM NEURAL NETWORK**

**V. YATSENKO, S. KRUCHININ, P. BIDYUK**

**Abstract.** The paper is focused on a new concept of a cryogenic-optical sensor intended for use in the space industry, geodynamics, and fundamental experiments. The basis of the sensor is a magnetic suspension with a levitating test body, a high-precision optical recorder of mechanical coordinates of the levitating body, and a signal-processing system. A Michelson-type interferometer with a laser diode and a single-mode optical fiber was used to measure the test body's displacements. The coordination of the laser diode coherence length and the difference in the interferometer optical lengths of the arms made it possible to eliminate coherent noise caused by interference from spurious reflections. The minimum recorded shift of the test body was 0.1 nm. The design of the sensor and the mathematical model of the superconducting suspension dynamics are presented. The results of experimental studies of a magnetic suspension together with an optical interferometric displacement sensor having a subnanometer sensitivity are shown.

**Keywords:** magnetic suspension, laser interferometer, optical fiber, displacement measurement, quantum neural network.

### **INTRODUCTION**

Remote sensing today is one of the rapidly growing modern measurement technologies. This direction of studies and applications is an industry with the cost of billions of dollars, and the number of distant specific images of different parts of the Earth is continuously growing. Solutions of many practical problems depend on the widespread use of measuring systems and different principles they use. These problems include monitoring of natural resources based upon analysis of gravitational anomalies, study of global geodynamic processes, the Earth's gravitational field, motions of Earth's poles, etc. To increase accuracy of the observations, determining the location and orientation for long-term air flights and underwater vehicle navigations require knowledge of Earth's gravitational field, including its anomalies. Detailed information on the Earth's gravitational field is needed by many industries and applied sciences (space research, geology, navigation, science of the Earth's shape). An accurate fast detection of geodynamic processes can provide data on the origin and development of critical local and global environmental conditions. Another practical problem is in the need to obtain more accurate information on undiscovered minerals of the Earth.

A gravimeter is a very accurate tool [1, 2, 6] for measuring the gravity acceleration  $g$ . At present, accuracy of the best stationary ground-based gravimeters is  $10^{-8} g$ . For sea-based gravimeters, it is  $10^{-7} g$ , and the aviation uses its value of about  $10^{-6} g$ . Most gravimeters manufactured by industry are based on the properties of a stretched spring or elastic properties of springs made of quartz or some special alloys. Their accuracy is not sufficient to solve these problems. Since the accuracy of gravimeters based on traditional principles has become fundamentally exhausted, many developers over the past decades have tried to use unconventional approaches in attempts to create ultraprecise gravimeters [1, 2]. These attempts can be grouped by the method of non-contact suspension of the gravimeter sensitive mass, by the use of electric or magnetic forces, by the methods of measuring the gravimeter sensitive mass displacements (optical recording systems, Josephson effects as the basis of measurements, etc.), as well as by computer based methods of signal processing. A great advance in the improvement of gravimeters was made due to the financing of development of superconducting gravimeters. J. M. Goodkind described in detail a superconducting gravimeter [1]. As he stated, the basic design of a superconducting gravimeter has been unchanged for almost 30 years since his first publication [2]. The free-state (levitation) of a sensitive mass of this gravimeter is achieved due to the Brownback-Meissner effect [3, 4]. An alternative approach is based on the phenomenon of magnetic levitation. The research in this field began in the late 1960s as a natural consequence of development of a low-temperature applied superconductivity, the theory of electromechanical conversion of energy and methods of control theory, as well as the development of defense navigation systems in the xUSSR. The phenomenon of “magnetic potential pit” (MPP) [4–6], discovered in 1975, means that the magnetic attraction between two distant magnets can change to the magnetic repulsion only as a result of growing distance between them. In this case the attractive force between two distant magnets was considered as a force that increases as the distance decreases. In other words, MPP considers the minimum magnetic potential energy as a function of the distance, although, by the time MPP was discovered, the magnetic interaction was considered monotonic, i.e. without the possibility of having a minimum anywhere except for points on the boundary.

The objective of this study is to present the results on the development of a sensitive element and a method for assessing the gravitational perturbation that affects a levitating test body. The paper presents the concept of a cryogenic optical sensor, its design, a mathematical model of dynamics of a superconducting suspension, and its stability analysis. The results of experimental studies of a magnetic suspension and an optical interferometric displacement sensor with subnanometer sensitivity are presented.

## **CONCEPT OF A CRYOGENIC OPTICAL SENSOR**

The sensor concept is based on three features and their combination [1, 2, 6, 7, 8]. First, this is a new type of superconducting suspension of a test body of highly sensitive gravimeter in a free state. Known superconducting suspensions [1–5]

use the concept of levitation based on the Brownback–Meissner effect [3], when the stable magnetic confinement without contact with other bodies arises due to the ideal diamagnetism of the superconductor, which ensures the expulsion of another magnet from its volume. As an example can be a superconducting magnet coil powered by the direct current. The test body of known gravimeters is, as a rule, continuous sphere. Instead of the superconductor diamagnetism, the new concept uses the zero resistance of a superconductor in the form of a thin closed loop, for example, a ring. Under certain conditions, due to the manifestation of the MPP phenomenon, such ring can steadily hang in a free state despite its diamagnetism that is practically not manifested and, according to well-known physical principles, makes the suspension inoperative. Another feature is the laser method for measuring the displacements of the test body, while known superconducting gravimeters use on the purpose sensors based on the Josephson effect (superconducting quantum interferometer sensors). The third feature is the use of modern signal processing methods to isolate very small disturbances against a background of significant noise levels that correspond to parameters of the measured gravitational field.

The choice of MPP as the basis for the levitation is explained by the two factors. The first of them is the desire to increase the sensitivity of the superconducting gravimeter, the second is to expand its dynamic range. The use of optical registration of the levitating sensitive mass displacements is also a new approach in the field of ultra-precise gravimeters. To process the signals of cryogenic optical gravimeter, it is proposed to use several processing steps. The first step is to compensate the noise affecting the base of the device. The second processing stage is focused on the use of the sensor inverse dynamic model [10]. The adaptive digital filtering is performed in the third stage of processing. In general, the novelty of the concept lies in the combination of a free sensitive mass suspension, an optical registration system, and new signal processing facilities. To our opinion, this approach to the sensor construction is implemented for the first time.

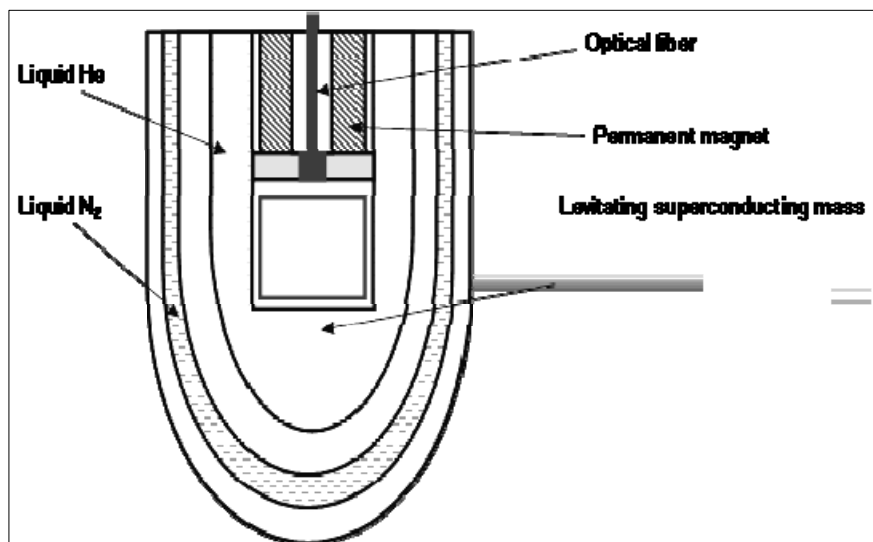
A superconducting gravimeter is a spring-type meter, in which, however, a magnetic returning force works as mechanical spring acting on a test superconducting body in an inhomogeneous magnetic field of superconducting rings or a permanent magnet. Due to the high stability of the superconducting currents, a highly stable non-dissipative spring is created. In equilibrium, the test body levitates in a position, where the gravitational force is balanced by a magnetic force acting in the opposite direction. When the gravity changes, the test body begins to shift from the zero position, and the electronic displacement sensor produces an error signal. By changing the current in the control ring, the auto-tuning system creates an additional magnetic field proportional to this signal, which keeps the test body in zero position. Since the returning force is proportional to the current, measuring the current in the control ring provides linear measurement of changes in the force of gravity.

In the model of sensitive element (Fig. 1), the test mass was in the form of a cylinder on which superconducting rings and a mirror for interferometric measurements were placed. Additional structural elements necessary for technological reasons were attached to this cylinder.

An effective use of the advantages of MSS (magnetic suspension system) which have almost unlimited sensitivity requires an appropriate system of regis-



tration of the test body displacement. To determine the test body positions in cryogenic element (CE), the use of a laser sensor was proposed. This made it possible to exclude probable disturbances in the position of test body caused by electric and magnetic fields as opposed to the conventional sensors used in previous systems. The modern interferometric methods and dynamic effects in the laser generation caused by weak external signals are used to detect ultra-small movements of the test body. The interferometric method can ensure the measuring accuracy of the test body coordinates not worse than 0.1 nm, what is sufficient to achieve the necessary sensor sensitivity. We have selected and implemented experimental schemes with laser displacement detectors, which provide measurements of the test body displacements and transformation of the signal into digital form for subsequent processing. It was shown that an optical sensor based on a diode laser with external resonator as a source of monochromatic radiation and a single-mode optical fiber as a channel for transporting light to a test body with maintaining the coherence of optical radiation that satisfies set requirements.



*Fig. 1. The layout of the opto-cryogenic sensor*

### **SENSOR DESIGN**

The suspension is coaxial, i.e. the holding magnets are displaced from the coaxial position to the position, where their axes are parallel to the axis of the suspension (Fig. 2). From various options for the number of holding magnets (two, three, four), a system of four rare earth permanent magnets with a vertical axis was selected. Each magnet in the horizontal plane has a rectangular shape. The vertical positions of all four sets of magnets were shifted from the axis of the suspension so that a space of 18 mm in diameter was formed to accommodate the optical sensor. The problem of the non-vertical position of the suspended free sample, which arose as a result of the unequal magnetic properties of the sets of permanent magnets, was solved by two structural changes. One of them was an increase in the pendulous feature of the test mass and the other was reduced to a

thin ferromagnetic ring, which compensated for the azimuthal inhomogeneity of the suspension magnetic field.

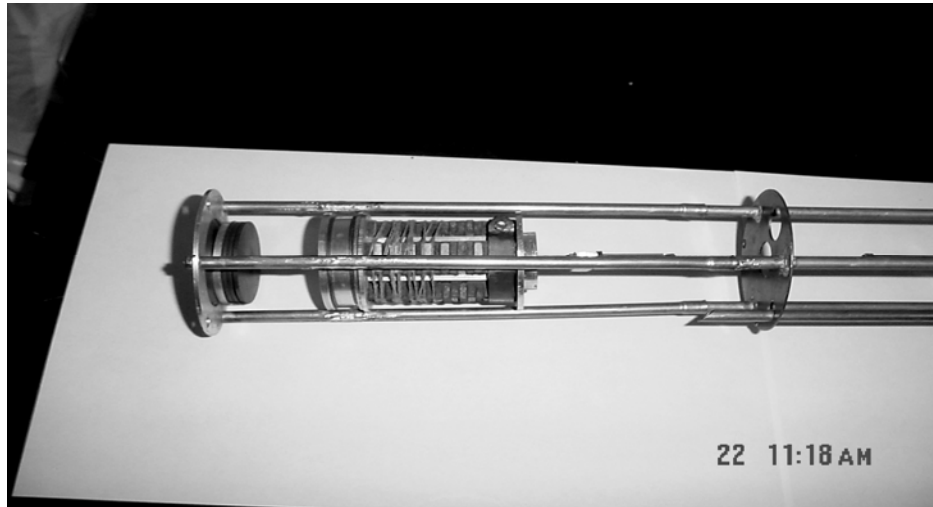


Fig. 2. Design of the opto-cryogenic sensor

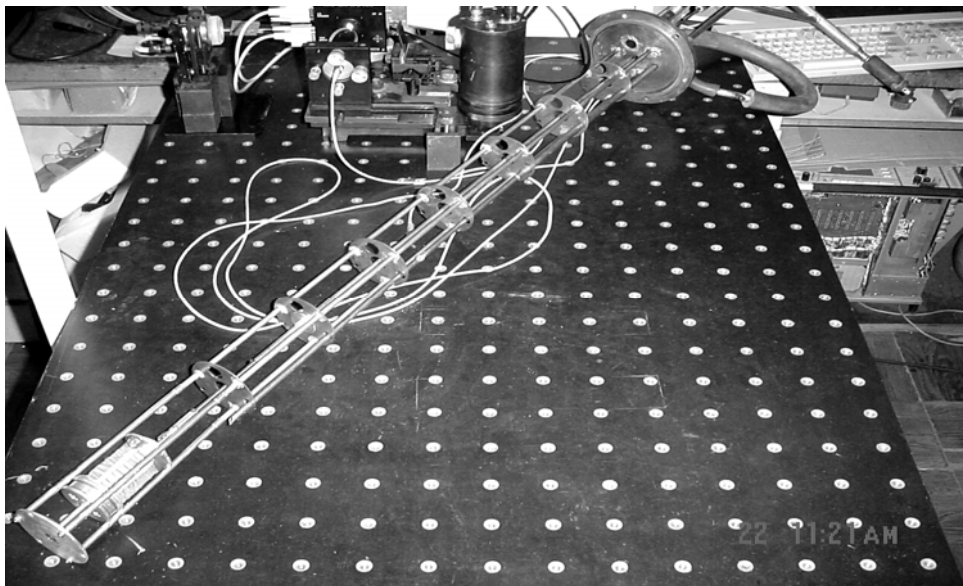
After theoretical and experimental studies of the suspension, optical sensor, and measurement software, the feasibility of the implementation of selected design as a whole was analyzed. As for the suspension, the main work concerned changes in the design of the magnetic system, when instead of placing remaining magnets on the axis of the suspension they had to be biased from the axis so that to place an optical laser sensor on it.

The new design of the working model (Fig. 3, 4) included four sets of permanent rare earth magnets, whose vertical axes of which were shifted from the suspension axis in four radial directions. The new design of the test mass had two niobium-titanium rings. The upper plane of the sample was polished as a reflective plane for laser beams. The levitation gap, depending on the weight of the test body, was from 7 to 15 mm. Based on this working model, theoretical and experimental studies of the joint operation of the suspension – registration system were carried out. The influence of the physical state of helium (liquid or gas) on the joint operation of the suspension – registration system was analyzed. The effect of a passive filter on the accuracy of the measurements was studied as well. The factors that influence a decrease in the suspension stiffness, in particular the presence of an additional ferromagnetic mass on a free sample were analyzed.

Experimental studies of the system were carried out to determine the properties of working model of a gravimeter, the dynamic characteristics of the magnetic suspension of free trial mass of the working layout of the gravimeter as well as the refinement of the sensitive element as a part of magnetic suspension, which was dictated by the experimental studies.



*Fig. 3. Magnetic suspension system of four sets of holding permanent magnets*



*Fig. 4. Magnetic suspension system and optical system for the measurement of mechanical coordinates*

### **SIGNAL PROCESSING SYSTEM**

The signal processing system consists of an adaptive compensator, inverse dynamic sensor model, adaptive Kalman filter, and a digital filter. Instead of a reverse model, a special type of neural network can be used. The software integrates gravity perturbation estimation algorithms with data processing models. It also includes a subsystem for interaction of the program core with the database, as well as with algorithms for its interaction with the sensor. It is supposed to include the software modules implemented in Matlab. Special tools are provided to

significantly reduce the levels of internal and light-generating noises based on the use of the features of the detector operation and software.

The analysis of the noise intensity characteristics of the optical coordinate meter was conducted. The correlation function and spectral density of the noise were obtained with regard for the properties of the bandpass filtering in an interferometer and the arbitrary modulation of a wavefront. The analysis included the modulation of a fractional noise as a special case of noises. To detect a signal limited by a fractional noise, the signal-noise ratio was found as a function of the modulation parameters. And the procedure for optimizing the noise-signal ratio corresponding to the signal demodulation was developed.

A noise compensation algorithm based on the global optimization approach is proposed. The compensator can extract useful information from a noisy optical signal. The noise compensation system allows the use of two types of sensors. The primary sensor generates a noisy source signal, whereas the secondary sensor measures the noise that is not correlated with the useful signal, but is correlated with a noise in the primary sensor.

Neural network algorithms for the signal analysis and estimation of gravitational perturbations based on the information approach have been developed. The error of the entropy minimization approach in identifying the dynamics of a test body was studied. An influence of the Parseval clock window on the search for a minimum of the entropy was studied. It has been analytically proved that the minimum of the entropy can be local. At the same time, the global minimum of the entropy with nonparametric estimation can be found by using information from the Shannon and Gaussian kernels. A comparative analysis of minimizing the entropy of the error and minimizing the entropy of the least squared error of a short-term prediction of experimental data is carried out. The statistical properties of the error in estimating the high-order central moments of the experimental time series and forecasting are used as comparison criteria. A mathematical description of a new structure of the inhibitor-type neural network which belongs to an important class of neural networks is developed. The necessary conditions for the behavior stability of a competitive inhibitory neural network are determined, and an algorithm based on stability conditions is developed. An algorithm of implementation of an inhibitory neural network for the evaluation of a signal characterizing the position of a levitating body in space has been also proposed.

Testing of the software for estimation of gravity perturbations affecting the levitating test body was performed. A module for the asymptotic estimation of gravitational perturbations by optical measurements has been added to the software. The analysis of the signal evaluation quality of the levitating body disturbance using the sequential processing of “optimal filter – inverse model – adaptive filter” was made. To compare the evaluation qualities, the following criteria were used: the ratio of standard deviations of the estimated signals, least-square error of the estimate, and the correlation coefficient of the signal without noise and the estimated signal.

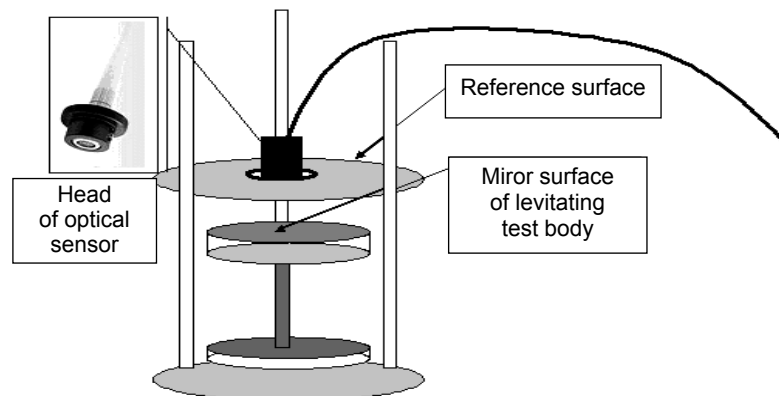
## **LASER DISPLACEMENT SENSOR OF A SUPERCONDUCTING GRAVIMETER**

Recent advances in development of diode lasers and the fiber optic technology, progress in the development of laser interferometry methods offer an alternative

to commonly used electronic displacement sensors. The use of fiberoptic interferometers has several important potential advantages [1], providing: 1) the possibility of creating a linear highly sensitive displacement detector; 2) absolute displacement measurements with a natural scale such as the laser radiation wavelength; 3) a dynamic range sufficient to record the largest seismic disturbances; 4) minimal electrical noise; 5) minimum size, relatively easy production, and the lack of electromagnetic and thermal noises; 6) the ability to reduce drifts while using frequency-stabilized lasers.

### **Optical Measurement System Design**

In the model of a superconducting gravimeter investigated in this paper, the motion of the test body was studied, by using an optical sensor — a laser fiberoptic interferometer. The laser interferometer is based on a 5-m long single-mode optical fiber, which made it possible to conveniently place the laser-optical unit relative to the cryogenic one. The test body had the shape of a cylinder on which a mirror was placed for interferometric measurements (Fig. 5).



*Fig. 5.* The scheme of the insert in a cryostat with an optical gage head mounted on a supporting plane and with the levitating test body

The general scheme of the optical measuring system is shown in Fig. 6. It included:

- a source of laser radiation;
- fiber optic cable;
- photodetector;
- signal recording system;
- laser radiation monitoring system.

The laser diode radiation was fed into the optical fiber using a collimating lens. The reference beam of the interferometer was formed by reflecting part of the laser radiation that was fed into the fiber from the output end of the light guide, and the signal beam was formed by the reflection from the polished aluminum surface of the test body. In order to reduce the sensitivity of the device to the angular deviations of the test body, the output collimating optics was not installed, the amplitude of the signal beam rapidly decreased with increasing the distance from the output end of the fiber to the test body. A normal operating range was considered at distances in the interval 0.2 – 0.8 mm.

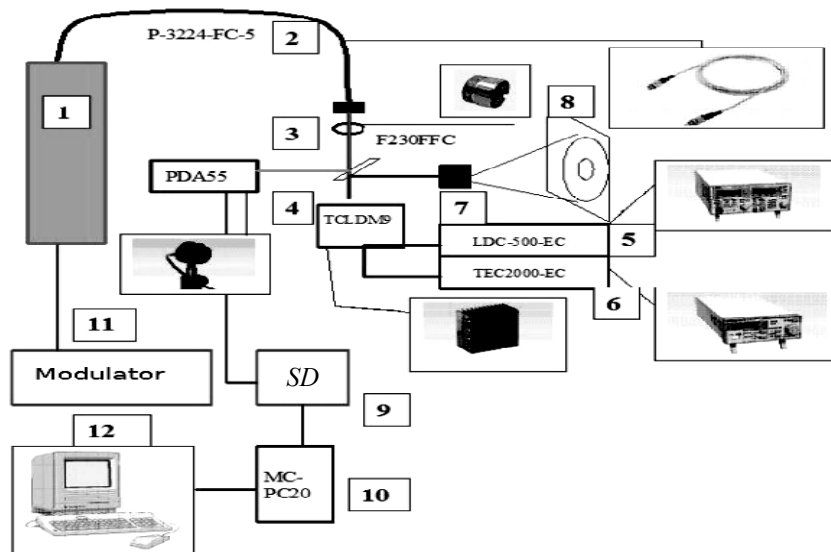


Fig. 6. Block diagram of the optical measuring system: 1 — cryostat with an insert on which an optical head is mounted; 2 — optical cable; 3 — focusing lens; 4 — a laser head with a laser diode; 5 — power supply of the laser diode; 6 — temperature stabilization unit; 7 — Fabry-Perot IFP-1 interferometer; 8 — screen; 9 — synchronous detector; 10 — analog-to-digital converter; 11 — modulator; 12 — computer

The reference and signal beams were returned through the fiber back to the laser optical unit, then reflected from the beam splitter and recorded by a photodetector.

The laser radiation source was based on a Hitachi semiconductor continuous laser HL6320G (emission wavelength 635 nm, output power 10 mW). The laser was placed in a TCLDM9 thermoelectric cooled laser head of Thorlabs Inc., which allowed the installation of diodes with a case diameter of 9 mm and 5.6 mm, allowing the modulation of a supply current in the frequency range from 100 kHz to 1 GHz and the precise control over the laser diode temperature. To feed the laser an LDC-500-EC electronic unit (Thorlabs Inc.) was used, which made it possible to set the laser diode current in the interval from zero to 500 mA with an accuracy of  $\pm 0.2$  mA. Noise in the frequency range of 10 Hz – 10 MHz does not exceed 5  $\mu$ A. The power supply had a mode for controlling the laser radiation power using the photodetector integrated in the laser diode case. The power supply current of the laser diode could be modulated using an external generator in the frequency range from DC to 150 kHz, the modulation coefficient is 50 mA / V.

The temperature of the laser diode was stabilized using a thermoelectric cooler built in a laser head, which was controlled by the TEC2000-EC thermal stabilization unit (Thorlabs Inc.). The temperature stability was 0.001 K, the adjustment accuracy was 0.01  $^{\circ}$ C.

The interferometer used a fiber optic cable P-3224-FC-5 (Thorlabs Inc.). The cable at both ends had ceramic caps, a numerical aperture of 0.12, a cut-off wavelength of 620 nm, is single-mode with a core diameter of 4  $\mu$ m, and a sheath diameter of 125  $\mu$ m. The cable length was 5 m. From the laser side, the radiation was fed into the optical fiber using an F230FFC-B collimating lens (Thorlabs

Inc.) with a diameter of 3.8 mm, a numerical aperture of 0.55, and a focal length of 4.5 mm. Optics were enlightened in the interval of 600–1050 nm.

As a photodetector, PDA55-EC silicon photodiode (Thorlabs Inc.) was used with a large receiving window area (13 mm<sup>2</sup>) with a low-noise amplifier, the gain of which was adjustable from  $1.5 \cdot 10^4$  V/A to  $1.5 \cdot 10^6$  V/A. With a minimum gain, the receiver bandwidth was at least 10 MHz.

The signal from the photodetector was either fed directly into an analog-digital 12-bit converter of MC-PC20 type or firstly into a synchronous detector to increase the signal-to-noise ratio, and then through an ADC to a personal computer.

To form the reference signal for the synchronous detector and to modulate the interference signal, we used a modulator, which generated an alternating high-voltage signal for supplying to the piezoelectric element to modulate the position of the reference plane of the interferometer. The recording system made it possible to record a continuous signal lasting 16s.

### **Properties of the fiber optic interference sensor and its maximum sensitivity**

**Interference signal.** The light intensity detected by a photodetector depends on the distance  $x$  between the test body and the end of the fiber and can be represented as

$$I = I_0 R_0 \left\{ R_1 + (1 - R_1)TR + 2\sqrt{T(1 - R_1)RR_1} \Gamma(\chi) \cos\left(4\pi \frac{x}{\lambda}\right) \right\}, \quad (1.1)$$

where  $I$  is the light intensity at the photodetector,  $I_0$  is the intensity of the laser radiation fed into the fiber,  $R_0$  is the reflection coefficient of the dividing plate,  $R_1$  is the reflection coefficient of the fiber end,  $R$  is the reflection coefficient of the surface of the test body,  $T$  is the collection efficiency of the fiber reflected from the surface of the test body,  $\Gamma(\chi)$  is the correlation function of laser radiation,  $\chi = \pi l / l_c$ , where  $l = 2x$  is the difference in the paths of the signal and reference rays,  $l_c$  is the coherence length of the laser radiation ( $l_c = x\Delta\nu$ , where  $c$  is the speed of light, and  $\Delta\nu$  is the width of the laser radiation spectrum). In the case of white frequency noise, the laser emission spectrum has a Lorentzian shape, and the correlation function has the form  $\Gamma(\chi) = \exp(-\chi)$ .

**The effect of laser radiation coherence.** Thus, the interference is possible, if the path difference is less than the coherence length. The coherence length substantially depends on the type of laser. For example, for a gas single-mode laser, it is  $l_c \approx 3$  km, while for laser diodes that were used in the described experiments, this value is less than 1cm. The relative smallness of the coherence length for laser diodes plays a very important role in the operation of a developed laser meter. In reality, in addition to the interference between the waves reflected from the test body and the end of the fiber, some of other interferences can be observed (e.g., the interference of waves reflected from the input and output ends of the fiber, etc.). This interference is an unwanted spurious effect that can significantly distort the useful signal. It is important that in the proposed scheme only a useful signal takes place, when the difference in the paths of the interfering waves is minimal, while, for all others, the difference in the courses contains the length of the fiber, which is 5 m in this experiment. Therefore, it can be expected that when using “bad” lasers with a short coherence length (but greater than the distance from the end of the fiber to the test body), the useful signal will not be distorted by the spu-

rious interference. At the same time, the use of “good” gas lasers will lead to a significant distortion of the signal.

To confirm this conclusion, an optical meter was experimentally tested for two cases where a He–Ne laser and a laser diode were used as a source. Fig. 7 illustrates the need to use low-coherence lasers. In this figure, the upper curve shows the interference signal, when using a gas laser, and the lower curve shows the interference signal for a laser diode.

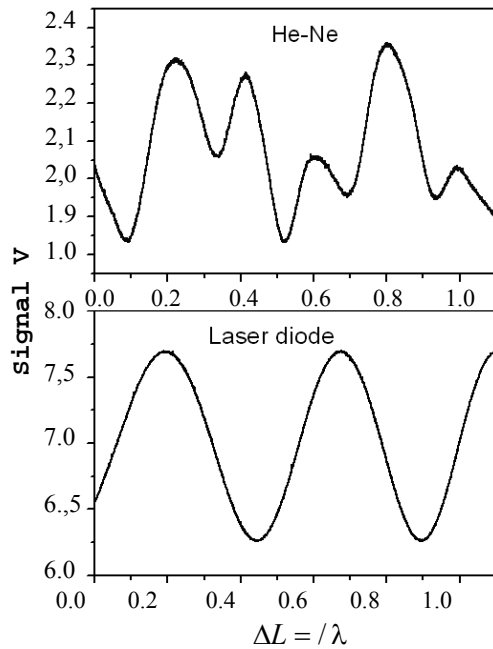


Fig. 7. Dependence of the photodetector signal on the displacement of the reflecting surface in the case of using a He–Ne laser (upper figure) and a laser diode (lower figure).

of the laser radiation wavelength ( $\lambda/2 \approx 0.32\mu\text{m}$ ) would be at a voltage of 48V. The sensitivity of the interferometer was determined when tuning the region of the highest steepness of the dependence of the signal on the displacement of the mirror.

It can be seen that, in the first case, the interference pattern is distorted due to spurious reflections, whereas, in the other case, the ideal sinusoidal dependence of the signal is observed, when the position of the test body changes.

**The sensitivity of the method.**

For experimental determination of the minimum displacement of a test body which can be recorded by the developed optical sensor, the scheme shown in Fig. 8 was used. The mirror simulating the test body was mounted on a piezoelectric element to which a constant voltage was applied to control the movement of the mirror with sub-nanometer accuracy. The magnitude of a mirror displacement was determined by the magnitude of the applied voltage, based on the fact that the displacement of the mirror by a half

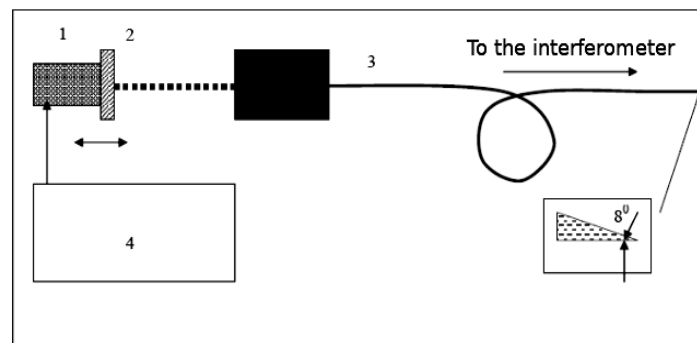


Fig. 8. Sensitivity calibration scheme of the optical sensor: 1 — piezoelectric element; 2 — mirror mounted on the piezoelectric element; 3 — optical sensor head; 4 — voltage source. The input end of the fiber is cut at an angle of  $8^\circ$



**Using a modulation technique to increase the sensitivity.** The analysis of the noise of the optical measuring system showed that, with an increase in the frequency of observations, the spectral density of the noise rapidly decreases, reaching a constant value at frequencies above 5kHz, which is two orders of magnitude lower than the spectral density at low frequencies. Therefore, to increase the accuracy of the optical sensor, the modulation technique can be used, which makes it possible to transfer the registration frequency to the spectral region, where the system noise is minimal.

According to (1.1), the dependence of the interference signal on the position of the mirror  $x$  can be written as

$$S = a + b \sin 2kx,$$

where  $k = 2\pi/\lambda$ ,  $a$  and  $b$  are constants. To eliminate the coordinate-independent component of the signal  $a$ , which makes the main contribution to low-frequency fluctuations, we modulate the distance to the test body with a frequency of  $\Omega$ ,  $x = x_0 + \Delta x \cos(\Omega t)$ .

The signal is recorded at the modulation frequency (or triple frequency) according to the formula

$$s_k = \int_0^{2\pi/\Omega} S(t) \cos(k\Omega t + \varphi) dt,$$

where  $k = 1, 3$ . After integration, we obtain, respectively:

$$s_1 = bJ_0(2k\Delta x) \cos\varphi \cos(2k\Delta x)$$

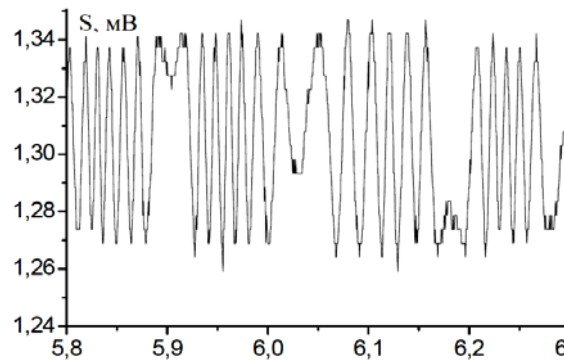
and

$$s_3 = b \cos\varphi \cos 2k\Delta x \left( \frac{2J_1(2k\Delta x)}{(k\Delta x)^2} - J_1(2k\Delta x) - \frac{2J_0(2k\Delta x)}{k\Delta x} \right),$$

where  $J_n(z)$  is the Bessel function.

Experimentally, the efficiency of the modulation technique was tested by changing the mirror position (or the position of the sensor optical head) with a frequency of 65Hz and with amplitude of about 0.1μm. The signal from the photodetector was fed to a selective amplifier, asynchronous detector, and an integrator and recorded by a computer. The registration of signals was performed

at a frequency of 65 Hz in a narrow frequency band (the averaging time was equal to 1 s), and the signal-to-noise ratio significantly increases. Fig. 7 shows the change in the signal with a stepwise displacement of the mirror with a step of 0.8nm. It can be seen that the use of the modulation technique even at a not very high modulation frequency leads to a significant reduction in the noise and, consequently, to a significant increase in the method



*Fig. 9. Registration of oscillations of a test body with an amplitude of about 1.0 μm and an oscillation period of about 0.25 s*

sensitivity. As follows from Fig. 9, the developed sensor allows to register displacements at the level of about 0.1 nm.

**Model experiments**

Given high cost of cryogenic experiments with relatively short levitation sessions in the gas phase, the development of the necessary technical solutions was carried out using models that recreate, as closely as possible, the real working conditions of a superconducting gravimeter. Based on the fact that the natural frequency of the developed test body was as low as several Hz, a scheme that recreates the dynamics of the test body by placing a reflecting mirror on a floating base was implemented.

It can be shown that the vertical oscillations of the body with an area of  $S$  that floats on the surface of a liquid with density  $\rho$  are described by the relation

$$m\ddot{x} = \rho g S x ,$$

where  $m$  is the mass of the float with a mirror,  $\rho$  is the density of a liquid, and  $g$  is the acceleration of gravity. The natural frequency of such a system is determined by the value  $\omega = \sqrt{\rho g S / m}$  and can easily be changed by choosing the appropriate values of the system parameters. For model experiments, a system was used with  $m = 0.05 \text{ kg}$ ,  $\rho = 10^3 \text{ kg / m}^3$  (water), the float had the shape of a cylinder with a diameter of 50 mm. The oscillation frequency of such a system was about 3 Hz.

Fig. 10 shows the signal from an optical sensor, when the oscillation amplitude was about 1 nm. It can be seen that, during the period of oscillations, the mir-

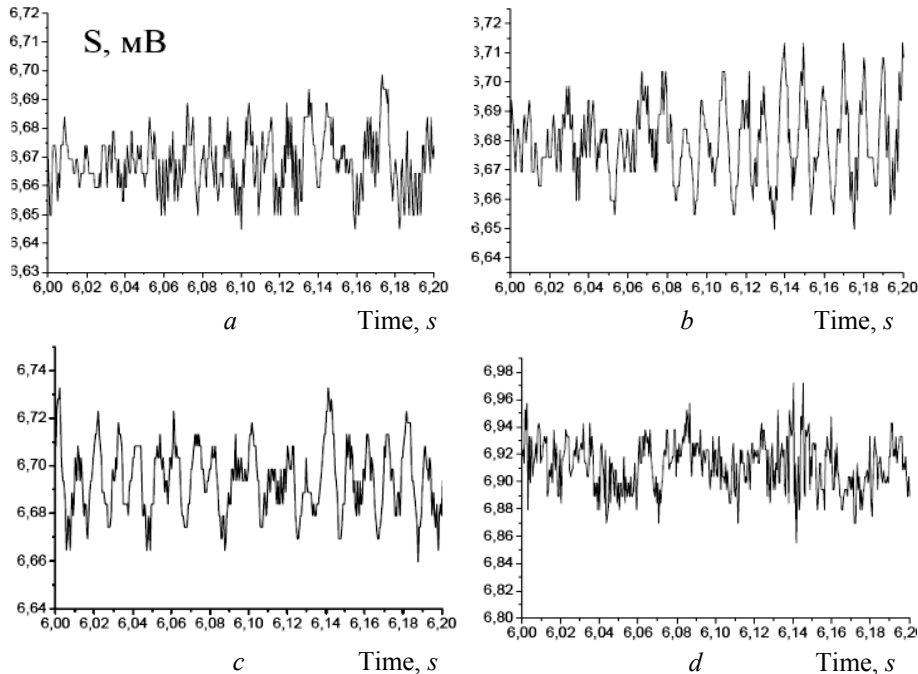


Fig. 10. An interference signal in the case of a levitating test body:  $a-c$  — fragments (0.2 s) of consecutive 16-second recordings of the interference signal, which were carried out after 3–5 min one by one. Levitating test body is in a gas environment without direct contact with liquid helium.  $d$  — the levitating test body is in direct contact with liquid helium

ror surface shifts by several half-waves, and the signal follows relation (1.1), in which the distance varies by the harmonic law.

The model experiments showed that the developed system makes it possible to register, after the appropriate mathematical processing of the obtained data, the vibrational motion of the test body with an amplitude greater than or comparable to the wavelength.

## **QUANTUM NEURAL NETWORKS**

It is planned to use in gravimeter a quantum neural network based upon quantum automates. We studied the quantum computing within the constraints of using a polylogarithmic ( $O(\log kn)$ ,  $k \setminus 1$ ) number of qubits and a polylogarithmic number of computation steps. The current research in the literature has focused on using a polynomial number of qubits. A new mathematical model of computation called Quantum Neural Networks (QNNs) is defined, building on Deutsch's model of quantum computational network. The model introduces a nonlinear and irreversible gate, similar to the speculative operator defined by Abrams and Lloyd. The precise dynamics of this operator are defined and while giving examples in which nonlinear Schrödinger's equations are applied, we speculate on its possible implementation. Many practical problems associated with the current model of quantum computing are alleviated in the new model. It is shown that QNNs of logarithmic size and constant depth have the same computational power as threshold circuits, which are used for modeling neural networks. QNNs of polylogarithmic size and polylogarithmic depth can solve the problems in NC, the class of problems with theoretically fast parallel solutions. Thus, the new model may indeed provide an approach for building scalable parallel computers.

Our quantum neural network based on quantum automatons. This useful possibility was introduced by V. Yatsenko. He used controllable Schrödinger's equations and it was shown how converts it to a quantum automaton. We formulate a new paradigm for computing with cellular automata (CAS) composed of arrays of quantum devices-quantum cellular automata. Computing in such a paradigm is edge driven. Input, output, and power are delivered at the edge of the CA array only; no direct flow of information or energy to internal cells is required. Computing in this paradigm is also computing with the ground state. The architecture is so designed that the ground-state configuration of the array, subject to boundary conditions determined by the input, yields the computational result. We propose a specific realization of these ideas using two-electron cells composed of quantum dots. The charge density in the cell is very highly polarized (aligned) along one of the two cell axes, suggestive of a two-state CA. The polarization of one cell induces a polarization in a neighboring cell through the Coulomb interaction in a very non-linear fashion. Quantum cellular automata can perform useful computing. The authors showed that AND gates, OR gates, and inverters can be constructed and interconnected. This opens new way for implementation of the gravimeter proposed.

## **CONCLUSIONS**

A gravimeter based on an optical sensor and a magnetic suspension of a superconducting test body was developed and experimentally investigated. The sensor

was studied at cryogenic temperatures, and procedures of data collection and processing were developed. Using the developed sensor, we studied the dynamics of a levitating test body in different environments (liquefied helium, cold helium vapor, etc.). On the studies of sensor's model its elements were refined to increase the sensitivity and to reduce the noise level.

An optical interferometric displacement sensor with subnanometer sensitivity was developed, created, and studied. The sensor was used to study the dynamics of oscillations of a test body with micron amplitudes, and it was shown that the proposed and experimentally implemented method for detecting the small displacements of the test body provides the possibility of using this method in superconducting gravimeters with adequate parameters of the sensitive element of a gravimeter. The achieved sensitivity of the optical sensor makes it possible to record a minimum displacement of the test body of the order of 100 pm, which with a natural frequency of oscillations of the test body of 0.2 Hz makes it possible to detect a change in the acceleration of gravity at a level of  $10^{10}$  g. For estimation makes the acceleration of gravity at a level of  $10^{-10}$  g we propose to use a quantum neural network based on quantum automatons.

## REFERENCES

1. W.A. Prothero and J.M. Goodkind, "A superconducting gravimeter," *Rev. Sci. Instr.*, vol. 39, pp. 1257–1261, 1968.
2. J.M. Goodkind and R.J. Warburton, "Superconductivity applied to gravimetry," *IEEE Trans. on Magn.*, vol. 11, iss. 2, 1975.
3. J.M. Goodkind, "The superconducting gravimeter," *Rev. Sci. Instrum.*, vol. 70, no. 11, pp. 4131–4152, 1999.
4. F.C. Moon, *Superconducting Levitation: Application to bearings and magnetic transportation*. NY: John Willey & Sons, 1994, 295 p.
5. S. Kruchinin, *Modern Aspect of Superconductivity: Theory of Superconductivity*. World Scientific, 2010, 232 p.
6. V. Kozoriz, *Novel Magnetic Levitation and Propulsion Phenomena*. Zaporizhyya, 1999, 271 p.
7. V. Yatsenko and P. Pardalos, "Global optimization of cryogenic-optical sensor," in *Sensors, Systems, and Next-Generation Satellites, Proc. SPIE*, 4550, pp. 433–441, 2001.
8. V. Yatsenko, "Functional structure of cryogenic optical sensor and mathematical modeling of signal," *SPIE Conference 'Optical Science and Technologies', 3-8 August 2003, San Diego, CA, USA, Proc. of SPIE*, vol. 5172.

Received 03.08.2022

## INFORMATION ON THE ARTICLE

**Vitaly O. Yatsenko**, ORCID: 0000-0002-7159-3312, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: gsaudr-sai@gmail.com

**Sergii P. Kruchinin**, ORCID: 0000-0002-0674-5826, Bogolyubov Institute of Theoretical Physics, NASU, Ukraine, e-mail: skruchin@i.com.ua

**Petro I. Bidiyuk**, ORCID: 0000-0002-7421-3565, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: pbidyuke\_00@ukr.net

**НАДПРОВІДНІ ГРАВІМЕТРИ НА ОСНОВІ СУЧАСНИХ НАНОМАТЕРІАЛІВ І КВАНТОВИХ НЕЙРОННИХ МЕРЕЖ** / В.О. Яценко, С.П. Кручинін, П.І. Бідюк

**Анотація.** Описано нову концепцію кріогенно-оптичного датчика, призначеного для використання у космічних дослідженнях, геодинаміці та фундаментальних експериментах. В основу датчика покладено магнітний підвіс з левітуючим тестовим тілом, високоточний оптичний реєстратор механічних координат левітуючого тіла і система оброблення сигналів. Як вимірювальну систему для визначення зміщень тестового тіла використано інтерферометр Міхельсона з лазерним діодом і оптоволоконном. Координація когерентної довжини лазерного діода і різниці оптичних довжин плечей інтерферометра дала змогу видалити когерентний шум, зумовлений інтерференцією від випадкових відбитків. Мінімально зареєстроване відхилення тестового тіла становило 0,1 нм. Подано процедуру проектування датчика, а також математичну модель динаміки надпровідної підвіски. Наведено результати експериментальних досліджень магнітної підвіски й оптичного інтерферометричного датчика відхилень, який має субнанометричну чутливість.

**Ключові слова:** магнітна підвіска, лазерний інтерферометр, оптичне волокно, вимірювання зміщення, квантова нейронна мережі.

## COMBINED CONTROL OF MULTIRATE IMPULSE PROCESSES IN A COGNITIVE MAP OF COVID-19 MORBIDITY

V. ROMANENKO, Y. MILIAVSKYI

**Abstract.** In this article, a cognitive map (CM) of COVID-19 morbidity in a given region was built. A general linear impulse process (IP) model in the CM was developed and measured, and unmeasured CM node coordinates were defined. The general IP model was decomposed into interrelated subsystems with measurable and unmeasurable node coordinates. For the subsystem with measurable node coordinates, multirate sampling of coordinates was conducted, resulting in the development of discrete dynamics models for quickly and slowly measured node coordinates. External controls were selected in IP models based on the possible variation of resources of node coordinates and CM weighting coefficients. IP control laws based on the variation of CM nodes and weight were designed. As a result, recurrent procedures for control generation in closed-loop control subsystems with multirate sampling were formulated. Experimental research on the control subsystems was carried out. It confirmed high efficiency for decreasing COVID-19 morbidity.

**Keywords:** cognitive map, impulse processes, control law, optimality criterion, COVID-19.

### INTRODUCTION

In the given article, cognitive modeling is applied to the research of dynamic processes of coronavirus morbidity. Cognitive modeling is based on the notion of a cognitive map (CM) which is defined as a weighted oriented graph with nodes representing coordinates (concepts, factors, characteristics) of a complex system and weighted edges (arcs) describing cause and effect interrelations between CM nodes. CM is built by experts. It allows qualitative description of interrelations between complex system's components and quantitative assessment of the effect of each CM node on all others, using edges of the oriented graph.

During evolution of a complex system with impulse-type behavior CM coordinates evolve with time under the effect of different disturbances. Each coordinate takes value  $z_i(k)$  at discrete time moments  $k = 0, 1, 2, \dots$ . At the next sampling period the value  $z_i(k+1)$  is determined by the value  $z_i(k)$  and information about increase or decrease of values of other nodes adjacent to given  $i$ -th node, at time moment  $k$ . Change of any  $j$ -th node at time moment  $k$  is called "impulse" and according to [1] is denoted by  $P_j(k)$  and is given as a difference

$P_j(k) = z_j(k) - z_j(k-1)$ ,  $k > 0$ . Impulse  $P_j(k)$  incoming to the  $j$ -th node will propagate over the paths of the CM to other nodes while increasing or decreasing. Propagation process of disturbances in the CM is defined by the difference equation [1]

$$z_i(k+1) = z_i(k) + \sum_{j=1}^n a_{ij} P_j(k), \quad i = 1, \dots, n, \quad (1)$$

where  $a_{ij}$  is a weight of an oriented graph's edge connecting the  $j$ -th node with the  $i$ -th one. If an edge connecting  $i$ -th and  $j$ -th nodes is absent, respective coefficient  $a_{ij} = 0$ .

CM nodes coordinates propagation rule (1) is often written as a first-order difference equation in variables increments

$$\Delta z_i(k+1) = \sum_{j=1}^n a_{ij} \Delta z_j(k), \quad (2)$$

which describes CM IP. Here  $\Delta z_i(k) = z_i(k) - z_i(k-1)$ ,  $i = 1, 2, \dots, n$ . In a vector form equation (2) is written as follows

$$\Delta \bar{Z}(k+1) = A \Delta \bar{Z}(k), \quad (3)$$

where  $A$  is a transposed adjacency matrix of the CM,  $\Delta \bar{Z}(k)$  is a vector of increments of coordinates  $z_i$  of CM nodes,  $i = 1, 2, \dots, n$ . From the control theory perspective the model (3) describes dynamics of linear multivariate system in discrete time under free motion of CM nodes.

Not all CM nodes are measurable in different complex systems. E.g., it is impossible to accurately measure level of a population health, level of democracy in a society, level of corruption and shadow economy, level of political activity etc. To solve this problem, [2] suggests to decompose the initial CM model (2), (3) into two interrelated CM. So,  $n$  CM nodes coordinates  $z_i$  are broken down into  $p$  measurable nodes  $x_i$  ( $i = 1, \dots, p$ ) and  $(n-p)$  unmeasurable nodes  $y_l$ ,  $l = p+1, \dots, n$ . Then IP model (3) can be presented as two interrelated subsystems of IP:

$$\Delta \bar{X}(k+1) = A_{11} \Delta \bar{X}(k) + A_{12} \Delta \bar{Y}(k); \quad (4)$$

$$\Delta \bar{Y}(k+1) = A_{21} \Delta \bar{Y}(k) + A_{22} \Delta \bar{X}(k), \quad (5)$$

where  $\bar{X}$  is a vector of measurable CM nodes,  $\bar{Y}$  is a vector of unmeasurable nodes. Here matrices of weights  $A_{12}$ ,  $A_{22}$  represent interrelations between the first (4) and the second (5) parts of the initial CM (3).

In [3] the problem of control automation for CM IP is solved by means of varying CM nodes coordinates and weights with unirate sampling, where controls are designed in a closed-loop control system. For this purpose equation (4) with measurable coordinates is used, augmented by controls as follows:

$$\Delta \bar{X}(k+1) = A_{11} \Delta \bar{X}(k) + A_{12} \Delta \bar{Y}(k) + B \Delta \bar{u}(k) + L(k) \Delta \bar{a}(k),$$

where  $\Delta \bar{u}(k) = \bar{u}(k) - \bar{u}(k-1)$  is the first difference of an external control vector formed by means of varying resources of CM nodes coordinates,

$\Delta\bar{a}(k) = \bar{a}(k) - \bar{a}(k - 1)$  is the first difference of a control vector based on varying the degree of influence  $a_{ij}(k)$  of the coordinate  $\Delta x_j(k)$  on the coordinate  $\Delta x_i(k)$ . The rules for writing matrices  $B$  and  $L(k)$  are described in [3, 4]. Applying variations  $\Delta\bar{u}(k)$  and  $\Delta\bar{a}(k)$  as control inputs is necessary when dimensions  $\dim\Delta\bar{u}(k)$  or  $\dim\Delta\bar{a}(k)$  are much less than number of nodes  $\dim\Delta\bar{X}(k)$ . In such a case using only one group of controls significantly decreases accuracy and speed of control systems for CM IP.

Among the coordinates of the vector  $\bar{X}$  in the model (4) there can be some coordinates  $\bar{X}_f$  measured (fixed) with a small sampling period  $T_0$  and some coordinates  $\bar{X}_s$  measured with a big sampling period  $h = mT_0$  where  $m$  is integer greater than 1. To describe dynamics of such a system a model of an IP with multirate sampling was developed in [5].

### CONSTRUCTION OF A COVID-19 MORBIDITY CM

Fig. 1 shows a CM of cause-and-effect relations in the process of spread of COVID-19 morbidity in a given region. The following nodes are included into the CM: 1 – number of daily revealed infected patients; 2 – number of daily vaccinated people; 3 – number of patients dying daily of COVID-19; 4 – number of

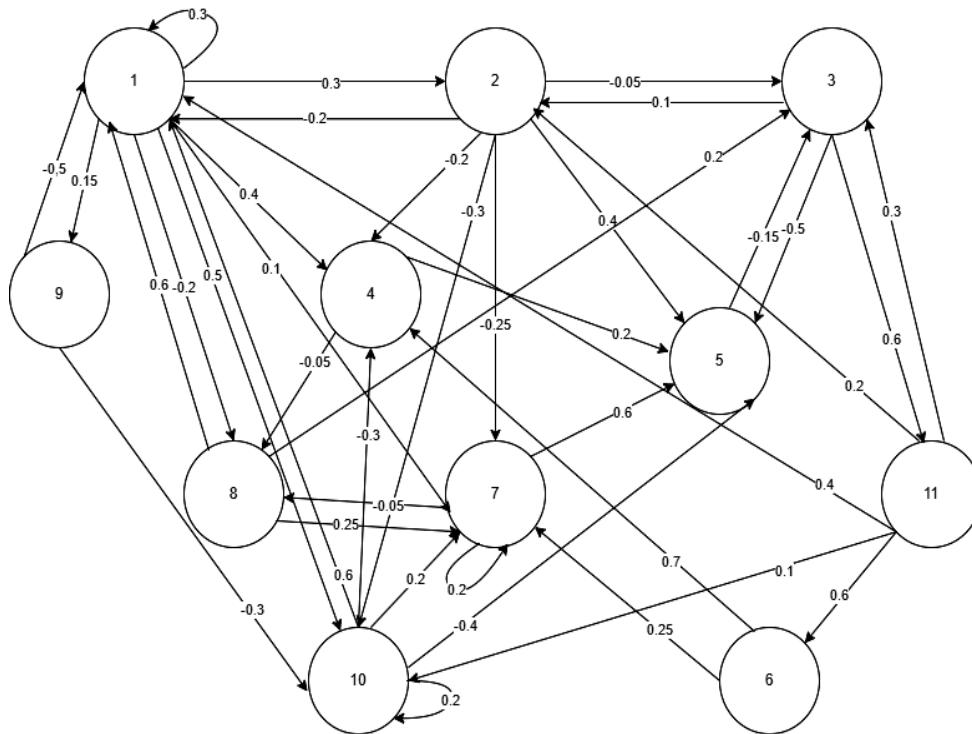


Fig. 1. COVID-19 morbidity CM

patients in isolation in the given region; 5 – number of patients recovered from COVID-19 in the given region; 6 – number of infected passengers revealed during arriving from other regions; 7 – number of patients in hospitals in the given



region; 8 – degree of contacts intensity set for the population of the given region when being in industrial, educational, public spaces; 9 – level of contact protection against infection (wearing masks); 10 – number of not isolated sick people who move freely (including those who arrived from other regions and are not revealed at arrival); 11 – level of danger of the virus strain.

Each CM node affects other nodes. Degree of the effect is estimated by the weights which can be positive or negative. E.g., the effect of the node 2 on the node 10 is reflected by the coefficient -0.3 because increase of number of daily vaccinated people leads to decrease of number of infected not isolated people who freely move; increase of level of danger of the virus strain (node 11) leads to increase of number of daily revealed infected patients with coefficient 0.4.

For cognitive modeling all CM nodes coordinates which represent factors of different physical nature are usually kept in a single scale. That's because when building the weighted oriented graph experts cannot correctly set the weighting coefficients of edges between CM nodes if they are measured in different units (like level of wearing masks, number of patients, danger of the virus strain). Here we suggest using 100 points scale for all CM nodes coordinates where 0 points means absence of a given factor and 100 points means maximal possible level of this factor at the given time interval. Obviously, when defining these factors values some subjectivity is possible, but it does not interfere with modeling or control of the whole system behavior.

#### **MODELS DEVELOPMENT FOR CM IP SUBSYSTEMS WITH MULTIRATE SAMPLING FOR QUICKLY AND SLOWLY MEASURED COORDINATES**

We develop models with multirate coordinates sampling based on the model (3), (5) of the subsystem with measurable coordinates. We assume that some coordinates  $\bar{X}_f$  of the vector  $\bar{X}$  belong to the quickly measured CM nodes with a small sampling period  $T_0$  and some coordinates  $\bar{X}_s$  are measured in discrete time moments with a big sampling period  $h = mT_0$ . Then the IP model (5) can be generally written in an intermediate form with unirate sampling as follows:

$$\begin{aligned} \begin{pmatrix} \Delta \bar{X}_f(k+1) \\ \Delta \bar{X}_s(k+1) \end{pmatrix} &= \begin{pmatrix} A_{11f} & A_{11s} \\ A_{11sf} & A_{11ss} \end{pmatrix} \begin{pmatrix} \Delta \bar{X}_f(k) \\ \Delta \bar{X}_s(k) \end{pmatrix} + \begin{pmatrix} B_f & 0 \\ 0 & B_s \end{pmatrix} \begin{pmatrix} \Delta \bar{u}_f(k) \\ \Delta \bar{u}_s(k) \end{pmatrix} + \\ &+ \begin{pmatrix} L_f(k) & 0 \\ 0 & L_s(k) \end{pmatrix} \begin{pmatrix} \Delta \bar{a}_f(k) \\ \Delta \bar{a}_s(k) \end{pmatrix} + \begin{pmatrix} A_{12f} \\ A_{12s} \end{pmatrix} \Delta \bar{Y}(k). \end{aligned} \quad (6)$$

Quickly measured coordinates  $\bar{X}_f$  are nodes 1, 2, 3, 4, 5, 6, 7 of the CM (Fig. 1). Slowly measured node  $x_s$  is the CM node 8 – degree of contacts intensity for the population of the given region. Unmeasured coordinates are nodes 9, 10, 11. To create controls  $\Delta u_2(k)$ ,  $\Delta u_4(k)$ ,  $\Delta u_8(k)$  we can use varying the resources of the nodes 2, 4, 8 respectively. Varying the weight coefficient  $a_{67}$  (how number of infected passengers revealed during arriving from other regions affects number of patients in hospitals) can be used to create a control  $\Delta a_{67}(k)$ . Then

sizes of the vectors in the model (6) will be  $\dim \bar{X}_f = 7$ ,  $\dim \bar{X}_s = 1$ ,  $\dim \bar{u}_f = 2$ ,  $\dim \bar{u}_s = 1$ ,  $\dim \bar{Y} = 3$ ,  $\dim \Delta \bar{a}_f = 1$ ,  $\dim \Delta \bar{a}_s = 0$ .

According to CM on Fig. 1, matrices in the model (6) will be the following:

$$A_{11f} = \begin{pmatrix} 0.3 & -0.2 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & -0.05 & 0 & 0 & -0.15 & 0 & 0 \\ 0.4 & -0.2 & 0 & 0 & 0 & 0.7 & 0 \\ 0 & 0.4 & -0.5 & 0.2 & 0 & 0 & 0.6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & -0.25 & 0 & 0 & 0 & 0.25 & 0.2 \end{pmatrix},$$

$$A_{11fs} = (0.6 \ 0 \ 0.2 \ 0 \ 0 \ 0 \ 0.25)^T,$$

$$A_{11sf} = (-0.2 \ 0 \ 0 \ -0.05 \ 0 \ 0 \ -0.05), \quad A_{11s} = 0.$$

Matrix  $B_f$  is created by the CM IP control system designer. It has to ensure scaling and switching designed controls  $\Delta \bar{u}_f(k)$ . Elements of the matrix  $B_f$  are zeros and ones. Element  $b_{i\mu} = 1$  when the  $i$ -th CM node is affected by the  $\mu$ -th component of the control vector. Thus in each row of the matrix  $B_f$  only one element can be equal to one and all others will be zero. Size of the matrix  $B_f$  for the given CM (6) is  $7 \times 2$  where 7 and 2 are sizes of vectors  $\bar{X}_f$  and  $\bar{u}_f(k)$  respectively. Then

$$B_f = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}^T.$$

The rules for writing the matrix  $L_f(k)$  can be found in [3]. For the model (6)  $L_f(k) = (0 \ 0 \ 0 \ 0 \ 0 \ \Delta x_6(k) \ 0)^T$ .

Unmeasurable CM nodes coordinates  $y_9, y_{10}, y_{11}$  in the model (6) are included into the vector  $\Delta \bar{Y}(k)$  as unmeasurable disturbances. Then matrix  $A_{12f}$  in the CM (Fig. 1) will be

$$A_{12f} = \begin{pmatrix} -0.5 & 0.6 & 0.4 \\ 0 & 0 & 0.2 \\ 0 & 0 & 0.3 \\ 0 & -0.3 & 0 \\ 0 & -0.4 & 0 \\ 0 & 0 & 0.6 \\ 0 & 0.2 & 0 \end{pmatrix}.$$

There is only one slowly measurable coordinate 8 in the CM which can be affected by the control  $u_s$  with a sampling period  $h = mT_0$  via varying resources of the node 8. So in model (6)  $B_s = 1, L_s(k) = 0$ . Unmeasurable nodes 9, 10, 11 don't affect the node 8, so  $A_{12s} = 0$ .

Thus IP model (5) is split into two parts. The first part describing dynamics of the quickly measured CM nodes with sampling period  $T_0$  can be written based on (6) as follows:

$$\begin{aligned} \Delta \bar{X}_f \left[ \left[ \frac{k}{m} \right] h + (l+1)T_0 \right] &= A_{11f} \Delta \bar{X}_f \left[ \left[ \frac{k}{m} \right] h + lT_0 \right] + A_{11fs} \Delta \tilde{x}_s \left[ \left[ \frac{k}{m} \right] h \right] + \\ &+ \left( B_f \quad L_f \left[ \left[ \frac{k}{m} \right] h + lT_0 \right] \right) \begin{pmatrix} \Delta \bar{u}_f \left[ \left[ \frac{k}{m} \right] h + lT_0 \right] \\ \Delta a_f \left[ \left[ \frac{k}{m} \right] h + lT_0 \right] \end{pmatrix} + \Delta \bar{\xi}_f \left[ \left[ \frac{k}{m} \right] h + lT_0 \right], \end{aligned} \quad (7)$$

where  $\left[ \frac{k}{m} \right]$  is integer part of dividing  $k$  by  $m$ ,  $l = 0, 1, \dots, m-1$ . The first difference

$$\begin{aligned} \Delta \bar{X}_f \left[ \left[ \frac{k}{m} \right] h + lT_0 \right] &= \bar{X}_f \left[ \left[ \frac{k}{m} \right] h + lT_0 \right] - \bar{X}_f \left[ \left[ \frac{k}{m} \right] h + (l-1)T_0 \right], \\ \Delta \tilde{x}_s \left[ \left[ \frac{k}{m} \right] h \right] &= x_s \left[ \left[ \frac{k}{m} \right] h \right] - x_s \left[ \left( \left[ \frac{k}{m} \right] - 1 \right) h \right] \text{ if } l = 0 \end{aligned}$$

and zero otherwise. Disturbances vector

$$\Delta \bar{\xi}_f \left[ \left[ \frac{k}{m} \right] h + lT_0 \right] = A_{12f} \Delta \bar{Y} \left[ \left[ \frac{k}{m} \right] h + lT_0 \right]$$

is generated by the unmeasurable nodes  $\Delta \bar{Y}$  of the CM.

The second part of the model described the dynamics of the slowly measured CM node 8 can be written in the intermediate form based on (6) as

$$\Delta x_s \left[ \left[ \frac{k}{m} \right] h + (l+1)T_0 \right] = A_{11sf} \Delta \bar{X}_f \left[ \left[ \frac{k}{m} \right] h + lT_0 \right] + B_s \Delta u_s \left[ \left[ \frac{k}{m} \right] h + lT_0 \right]. \quad (8)$$

Based on the iterative procedure described in [5] the model (8) can be transitioned to the form where the coordinate  $x_s$  and the control  $u_s$  have the big sampling period  $h = mT_0$ , considering that  $A_{11s} = 0$  and there are no external disturbances and weights-varying based controls:

$$\Delta x_s \left[ \left( \left[ \frac{k}{m} \right] + 1 \right) h \right] = A_{11sf} \Delta \tilde{X}_f \left[ \left[ \frac{k}{m} \right] h + (m-1)T_0 \right] + B_s \Delta u_s \left[ \left[ \frac{k}{m} \right] h \right], \quad (9)$$

where 
$$\Delta x_s \left[ \left( \left[ \frac{k}{m} \right] + 1 \right) h \right] = x_s \left[ \left( \left[ \frac{k}{m} \right] + 1 \right) h \right] - x_s \left[ \left[ \frac{k}{m} \right] h \right];$$

$$\Delta \tilde{X}_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + (m-1)T_0 \right] = \bar{X}_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + (m-1)T_0 \right] - \bar{X}_f \left[ \left( \begin{bmatrix} k \\ m \end{bmatrix} - 1 \right) h + (m-1)T_0 \right].$$

### CONTROL AUTOMATION OF CM IP WITH MULTIRATE SAMPLING

For designing algorithms of CM IP automated control dynamics (7), (9) of the vectors  $\bar{X}_f$ ,  $x_s$  should be written in full CM nodes coordinates (not in increments):

$$\begin{aligned} & \bar{X}_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + (l+1)T_0 \right] = \\ & = (I_{11f} + A_{11f} - A_{11f}q^{-1})\bar{X}_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \right] + A_{11fs}\Delta \tilde{x}_s \left[ \begin{bmatrix} k \\ m \end{bmatrix} h \right] + \\ & + \left( B_f \quad L_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \right] \right) \begin{pmatrix} \Delta \bar{u}_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \right] \\ \Delta a_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \right] \end{pmatrix} + \Delta \bar{\xi}_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \right]; \quad (10) \\ & x_s \left[ \left( \begin{bmatrix} k \\ m \end{bmatrix} + 1 \right) h \right] = \\ & = x_s \left[ \begin{bmatrix} k \\ m \end{bmatrix} h \right] + A_{11sf}\Delta \tilde{X}_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + (m-1)T_0 \right] + B_s \Delta u_s \left[ \begin{bmatrix} k \\ m \end{bmatrix} h \right], \quad (11) \end{aligned}$$

where  $q^{-1}$  is a reverse shift operator with sampling period  $T_0$ .

To design quickly changes controls  $\Delta \bar{u}_f, \Delta a_f$  the following quadratic criterion is suggested:

$$\begin{aligned} J_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + (l+1)T_0 \right] &= E \left\{ \left[ \bar{X}_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + (l+1)T_0 \right] - \bar{G}_f \right]^T \times \right. \\ & \times \left[ \bar{X}_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + (l+1)T_0 \right] - \bar{G}_f \right] + \left( \Delta \bar{u}_f^T \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \right] \quad \Delta a_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \right] \right) \times \\ & \left. \times \begin{pmatrix} R_{f1} & 0 \\ 0 & r_{f2} \end{pmatrix} \begin{pmatrix} \Delta \bar{u}_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \right] \\ \Delta a_f \left[ \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \right] \end{pmatrix} \right\}, \quad (12) \end{aligned}$$

where  $E$  is expectation (mean),  $\bar{G}_f$  is set-point vector for stabilization of CM nodes coordinates  $\bar{X}_f$ .

Based on minimization of criterion (12) with respect to vector  $\begin{pmatrix} \Delta \bar{u}_f \\ \Delta a_f \end{pmatrix}$ , having used model (10), we find quick combined control vector that affects nodes  $\bar{X}_f$  according to (10):

$$\begin{aligned} & \begin{pmatrix} \Delta \bar{u}_f \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \\ \Delta a_f \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \end{pmatrix} = \\ & = - \left\{ \begin{pmatrix} B_f^T \\ L_f^T \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \end{pmatrix} \begin{pmatrix} B_f & L_f \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \end{pmatrix} + \begin{pmatrix} R_{f1} & 0 \\ 0 & r_{f2} \end{pmatrix} \right\}^{-1} \begin{pmatrix} B_f^T \\ L_f^T \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 \end{pmatrix} \times \\ & \quad \times \left\{ (I_{11f} + A_{11f} - A_{11f}q^{-1}) \bar{X}_f \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 + A_{11fs} \Delta \tilde{x}_s \begin{bmatrix} k \\ m \end{bmatrix} h + \right. \\ & \quad \left. + \Delta \bar{\xi}_f \begin{bmatrix} k \\ m \end{bmatrix} h + lT_0 - \bar{G}_f \right\}. \end{aligned} \quad (13)$$

To design the slow control  $\Delta u_s$ , the second optimality criterion is suggested:

$$J_s \left[ \left( \begin{bmatrix} k \\ m \end{bmatrix} + 1 \right) h \right] = E \left\{ \left[ x_s \left( \begin{bmatrix} k \\ m \end{bmatrix} + 1 \right) h - G_s \right]^2 + r_s \Delta u_s^2 \begin{bmatrix} k \\ m \end{bmatrix} h \right\}, \quad (14)$$

where  $G_s$  is set-point vector for  $x_s$  stabilization. Based on minimization of criterion (14) with respect to  $\Delta u_s$ , having used model (11), we find slow combined control that affects node 8:

$$\Delta u_s \begin{bmatrix} k \\ m \end{bmatrix} h = - \frac{B_s}{B_s^2 + r_s} \left( x_s \begin{bmatrix} k \\ m \end{bmatrix} h + A_{11sf} \Delta \tilde{X}_f \begin{bmatrix} k \\ m \end{bmatrix} h + (m-1)T_0 - G_s \right). \quad (15)$$

## EXPERIMENTAL RESEARCH OF THE IP CONTROL SYSTEM IN THE COVID-19 MORBIDITY CM

For a computational simulation initial values of CM nodes coordinates were set at the medium levels  $x_i = 50$ ,  $i = 1, \dots, 11$ . Problem statement of the experiments is to move CM nodes coordinates  $x_2, x_4, x_7, x_8$  to the new levels  $G_2 = 60$ ,  $G_4 = 40$ ,  $G_7 = 40$ ,  $G_8 = 40$ . It means that we need to increase number of daily vaccinated people ( $x_2$ ), decrease number of patients in isolation in the given region ( $x_4$ ), decrease of patients in hospitals in the given region ( $x_7$ ), decrease degree of contacts intensity set for the population of the given region when being in industrial,

educational, public spaces ( $x_8$ ). In the fast control subsystem (13) control vector consists of the controls  $\Delta \bar{u}_f = (\Delta u_2 \ \Delta u_4)^T$  and  $\Delta a_f = \Delta a_{67}$ , and in the slow subsystem (15) there is only one control  $\Delta u_s(rh) = \Delta u_8(rh)$ , where  $r = \left[ \frac{k}{m} \right]$ . Ratio between sampling periods of fast and slow subsystems  $h = mT_0$  is selected with the coefficient  $m = 6$ .

Fig. 2 shows the charts for the results of simulation of CM nodes  $x_i, i=1, \dots, 11$ , and Fig. 3 demonstrates charts of the generated increments of controls  $\Delta u_2(kT_0), \Delta u_4(kT_0), \Delta a_{67}(kT_0), \Delta u_8(rh)$  based on control laws (13) and (15).

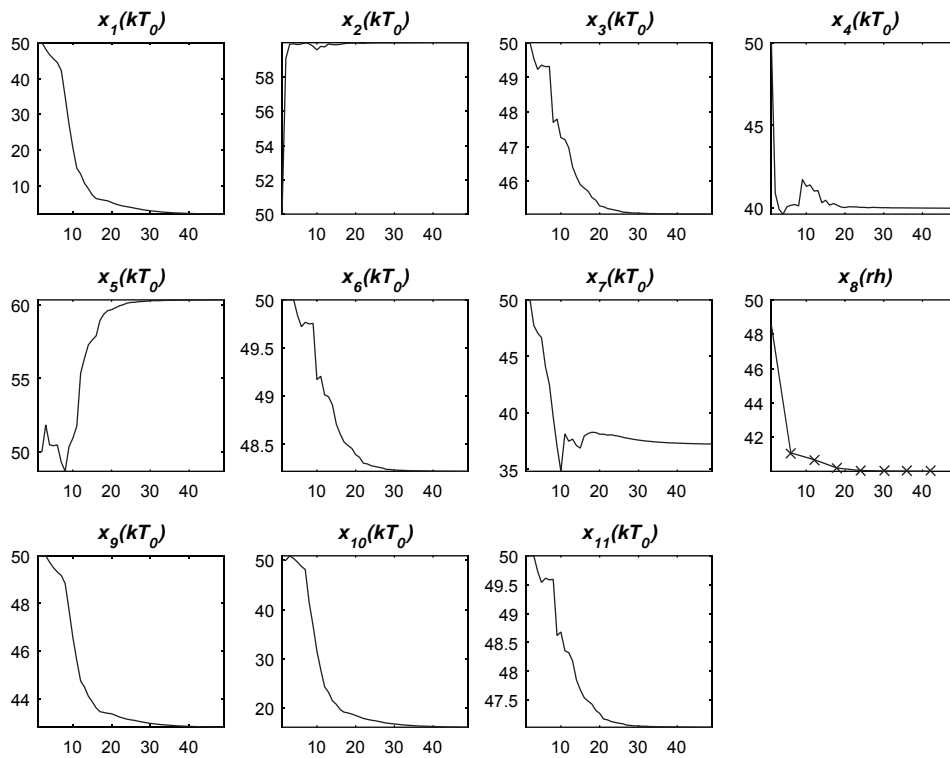


Fig. 2. Nodes coordinates changes

Based on the charts analysis we can formulate the following tendencies in the changes of CM nodes coordinates and controls with multirate sampling.

1. CM nodes coordinates  $x_2, x_4, x_7, x_8$  which are directly controlled by  $\Delta u_2(kT_0), \Delta u_4(kT_0), \Delta a_{67}(kT_0), \Delta u_8(rh)$  respectively quickly shift to their new levels  $G_2 = 60, G_4 = 40, G_7 = 40, G_8 = 40$ .

2. Nodes coordinates  $x_1, x_3, x_5, x_6$  which are not controlled directly move slower to the new natural level, i.e.  $x_1, x_3, x_6$  decrease and  $x_5$  increases.

3. Controls in the form of increments  $\Delta u_2(kT_0), \Delta u_4(kT_0), \Delta a_{67}(kT_0), \Delta u_8(rh)$  set at zero levels when transient processes are over.

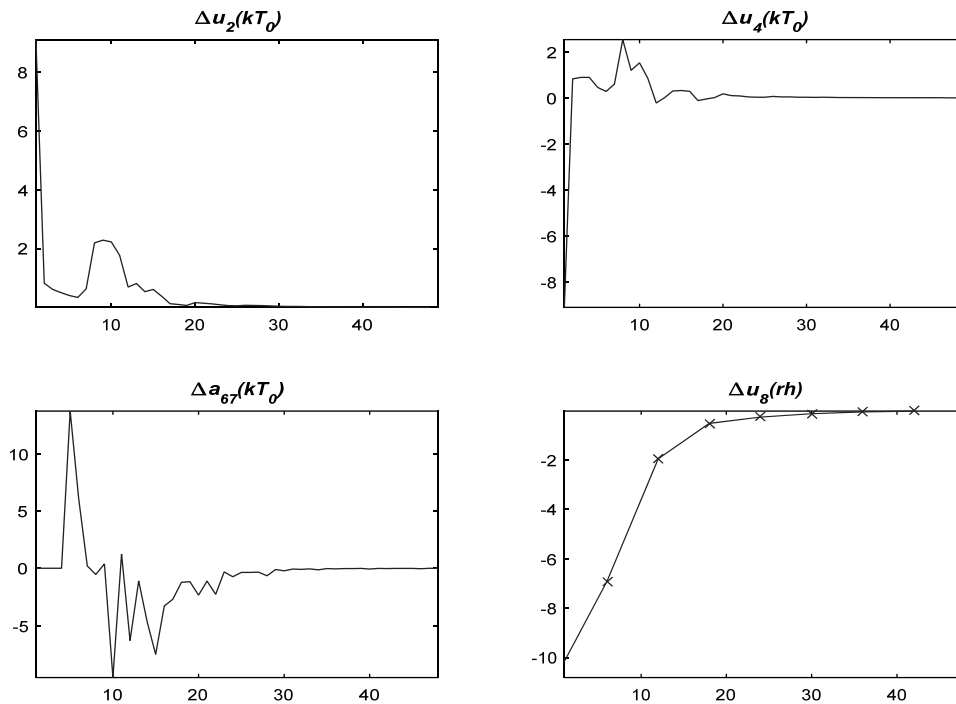


Fig. 3. Controls changes

## CONCLUSIONS

This article develops a CM which quantitatively describes interrelations between factors during the spread of COVID-19 morbidity. Based on the CM the IP models are developed which describe dynamic aspects of the morbidity in the form of difference equations. Both measurable and unmeasurable coordinates are accounted for. We also account for presence of both quickly and slowly measured coordinates, which are reflected in the subsystems models with fast and slow sampling. For these subsystems external controls are selected, which are then generated based on the linear quadratic control method using combined resources varying of some nodes coordinates and edges weights in the CM.

Experimental research was conducted by means of computational simulation of the CM IP closed-loop control system. Based on the charts of the transients processes of CM nodes coordinates and incremental controls it is concluded that directly controlled CM nodes quickly shift to the new levels defined by the set-points of the controller. Not directly controlled CM nodes coordinates move to the new levels slower.

## REFERENCES

1. F. Roberts, *Discrete Mathematical Models with Applications to Social, Biological, and Environmental Problems*. Englewood Cliffs, Prentice-Hall, 1976, 559 p.
2. Mikhail Z. Zgurowsky, Victor D. Romanenko, and Yuriy L. Milyavskiy, "Principles and Methods of Impulse Processes Control in Cognitive Maps of Complex Systems.

- Part 1,” *Journal of Automation and Information Sciences*, vol. 48, no. 3, pp. 36–45, 2016. doi: 10.1615/JAutomatInfScien.v48.i3.40.
3. V. Romanenko and Y. Milyavsky, “Control automation method in cognitive maps based on the synthesis of increments of weighting coefficients and nodes coordinates,” (in rus.), *System Research and Information Technologies*, no. 3, pp. 89–99, 2019. doi: 10.20535/SRIT.2308-8893.2019.3.08.
  4. V. Romanenko and Y. Milyavsky, “Control method in cognitive maps based on weights increments,” *Cybernetics and Computer Engineering*, issue 184, pp. 44–55, 2016. doi: 10.15407/kvt184.02.044.
  5. V. Romanenko, Y. Miliavskyi, and H. Kantsedal, “Application of Impulse Process Models with Multirate Sampling in Cognitive Maps of Cryptocurrency for Dynamic Decision Making,” in *System Analysis & Intelligent Computing. Theory and Applications*; eds: M. Zgurovsky, N. Pankratova. Springer, 2022, pp. 115–137. doi: 10.1007/978-3-030-94910-5.

Received 10.08.2022

### INFORMATION ON THE ARTICLE

**Victor D. Romanenko**, ORCID: 0000-0002-6222-3336, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: romanenko.viktorroman@gmail.com

**Yurii L. Miliavskyi**, ORCID: 0000-0003-0882-3418, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: yuriy.milyavsky@gmail.com

### КОМБІНОВАНЕ КЕРУВАННЯ ІМПУЛЬСНИМИ ПРОЦЕСАМИ З РІЗНОТЕМПОВОЮ ДИСКРЕТИЗАЦІЄЮ В КОГНІТИВНІЙ КАРТІ ЗАХВОРЮВАНOSTІ НА COVID-19 / В.Д. Романенко, Ю.Л. Мілявський

**Анотація.** Побудовано когнітивну карту (КК) розповсюдження захворюваності на COVID-19 в даному регіоні. Розроблено загальну лінійну модель імпульсних процесів (ІП) КК і проведено аналіз вимірюваних і невимірюваних координат вершин КК. Виконано декомпозицію загальної моделі ІП на взаємопов’язані підсистеми з вимірюваними і не вимірюваними координатами вершин. Для підсистеми з вимірюваними координатами вершин проведено різнометрову дискретизацію координат, у результаті чого розроблено дискретні моделі динаміки для швидковимірюваних і повільновимірюваних координат вершин КК. Вибрано зовнішні керувальні дії в моделях ІП з урахуванням можливого варіювання ресурсами координат вершин і вагових коефіцієнтів ребер КК. Виконано синтез законів керування ІП на основі варіювання координат вершин і вагового коефіцієнта. Розроблено рекурентні процедури формування керувальних дій у замкнених підсистемах керування з різнометровою дискретизацією. Проведено експериментальні дослідження підсистем керування, які підтверджують високу ефективність по зниженню захворюваності на COVID-19.

**Ключові слова:** когнітивна карта, імпульсні процеси, закони керування, критерій оптимальності, COVID-19.



## THE USE OF ENVIRONMENTAL DECISION SUPPORT SYSTEMS FOR MODELING OF ATMOSPHERIC POLLUTION FOLLOWING THE CHEMICAL ACCIDENTS

I.V. KOVALETS, V.P. BESPALOV, S.Ya. MAISTRENKO, O.I. UDOVENKO

**Abstract.** We studied the possibility of the combined application of screening models to assess the characteristics of sources in accidents at storage facilities for hazardous substances with complex models of atmospheric transport as part of modern decision support systems to calculate air pollution in a wide range of spatial and temporal scales. The evaporation time following an emergency spill, estimated by screening models, is used to set the emission intensity and calculate the atmospheric transport by the RODOS nuclear emergency response system. For the accident in Chernihiv on March 23, 2022, it was estimated that the maximum permissible concentration of ammonia 0.2 mg/m<sup>3</sup> was exceeded at distances up to 75 km from the source. The dependence of the calculated maximum concentrations on time is close to asymptote  $c_{\max} \sim t^{-4.5}$  up to 15 h after emission, which is consistent with the asymptote  $\sigma \sim t^{2/3}$  for the time dependence of the sizes of puffs following turbulent dispersion of instantaneous releases.

**Keywords:** atmospheric dispersion, RODOS system, hazardous substances, ammonia, LASAT, DIPCOT, “Povitrya” system.

### INTRODUCTION

The military aggression of Russia has led, in particular, to numerous cases of man-made accidents in Ukraine with the emission of potentially hazardous substances (HS) into the atmosphere. In many cases, hazardous chemicals such as ammonia, chlorine, and others are stored in liquefied form. As a result of damage to a tank or pipeline, a rapid release of the gaseous phase into the atmosphere (primary cloud) and a spill of the liquid phase of the corresponding chemicals onto the underlying surface with subsequent evaporation (secondary cloud) can occur simultaneously. Operational forecasting of damage zones as a result of the spread of clouds formed after the spill of HS in the world and in Ukraine is carried out according to simplified methods, which are called “screening models” [1]. Such models describe both the formation of an emission source through the spread and evaporation of a spill and the spread of a cloud of HS. Examples of such screening models are works [1–4]. In many cases, the initial results of screening models are only the integral characteristics of the cloud of HS, such as the radius of the affected area. When propagating at a distance of more than 10–20 km, it is necessary to take into account the influence of spatially and temporally variable meteorological conditions on the formation of pollution concentration fields, which is possible by using state-of-art atmospheric transport models (ATMs) that simulate cloud propagation based on data from numerical weather forecast models and are able to calculate atmospheric transport at distances from

~0.1 to ~1000 km. Examples of modern ATMs are the LASAT [5] and DIPCOT [6] models as part of the EU nuclear emergency response system RODOS [7], the WRF-CHEM and CMAQ Euler models [8], the FLEXPART Lagrangian model, and the CALPUFF Lagrangian-Eulerian model [9] and others. All of the above models are widely used in the world both for forecasting pollution in real-time and for analyzing environmental risks associated with possible emissions of pollutants into the atmosphere. In a number of cases, intermediate results of screening model calculations can be used to set the source characteristics in atmospheric transport models for further modeling the distribution of a cloud of HS. The purpose of this work is to study the possibility of combined use of screening models for assessing the characteristics of emission sources during accidents at hazardous chemicals storage and transportation facilities with the state-of-art atmospheric transport models for calculating atmospheric pollution over a wide range of spatial and temporal scales.

## **METHODS**

In this work, we use the atmospheric transport model of the RODOS system of the European Union for nuclear emergency response [7]. It was recently implemented in Ukraine to predict the consequences of radiation accidents and analyze the risks associated with possible accidents [10, 11]. The RODOS system uses several atmospheric transport models. In this work, we used the LASAT model of Lagrangian particles [5], which is the most resource-demanding, but at the same time the most suitable for application in the case of emissions near the Earth's surface. In the LASAT model, particles are considered idealized material points, which are not assigned a spatial size. Other atmospheric models of the RODOS system use so-called "puffs" — instantaneous emissions that represent long-term emission. Each puff is assigned a Gaussian distribution of matter in space, and the vertical and horizontal size of the puff is determined by the dispersions of this distribution in the horizontal and vertical planes  $\sigma_h^2$ ,  $\sigma_v^2$ . The distribution dispersions increase with time, but when propagating in flows with a large velocity shift, for example, near the Earth's surface, the symmetry of the distribution in real clouds is violated, which may be not well described by the corresponding models.

The input meteorological data of the RODOS system in Ukraine are the results of numerical weather forecasting of the WRF-Ukraine system, which operates at the Ukrainian Hydrometeorological Center and uses the modern WRF meteorological model for calculations [12]. The weather prediction data of the WRF-Ukraine system have been repeatedly verified on the basis of hydrometeorological measurements in Ukraine [13].

The source parameters were calculated using the "Povitrya" system [14], which is in the public domain [15]. The system "Povitrya" implements screening models for predicting damage zones during HS emissions [3, 4], the intermediate results of which are the mass of the primary cloud ( $m_1$ ) and the evaporation time ( $\tau$ ) of the secondary cloud. The mass of the primary cloud depends, in particular, on the boiling point of the given substance, the heat of boiling, the specific heat of the substance, and the air temperature. Since the total mass of matter in the container  $m_0$  is set by the user, the emission intensity  $q_s$  during the formation of a secondary cloud by evaporation can obviously be calculated by the formula:

$$q_x = \frac{m_2}{\tau} = \frac{m_0 - m_1}{\tau}. \quad (1)$$

Here  $m_2$  is the mass of the secondary cloud. For most substances in the model [3], the mass of the primary cloud is neglected, and then  $m_2 = m_0$ . In [16], examples of the application of Ukrainian and foreign models for calculating the evaporation rate  $q_s$  and the mass of the primary cloud and their comparison are given. It is shown that Ukrainian [3, 4] and foreign [2] models give comparable results.

Fig. 1 presents a flowchart for the calculation of air pollution based on the combined use of screening models to assess the characteristics of sources of emissions of hazardous chemicals with state-of-art models of atmospheric transport. For the calculations of the screening model, it is necessary to set a relatively small number of input parameters: the volume of the spill, the conditions of the spill (was there a spill containment pallet or not), the storage conditions of the substance (under pressure or not), the type of terrain (for example, “flat terrain”), the type of vegetation (for example “grassland”), meteorological conditions at the time of the spill (wind speed and direction, surface air temperature, soil temperature if available, atmospheric stability). The masses of primary and secondary clouds calculated on the basis of the screening model, as well as the evaporation time, are used to set the emission intensity in the atmospheric transport model according to the formula (1). The mass of the primary cloud is accounted for as an instantaneous or short-term emission. The calculated characteristics of the source are transferred to the ATM, which calculates the spread of atmospheric pollutants taking into account meteorological conditions changing in time and space. In this work, the transfer of the estimated characteristics of the emission source to the ATM was performed manually. The result of the ATM calculation is the field of near-surface concentration of the pollutant distributed in time and space.

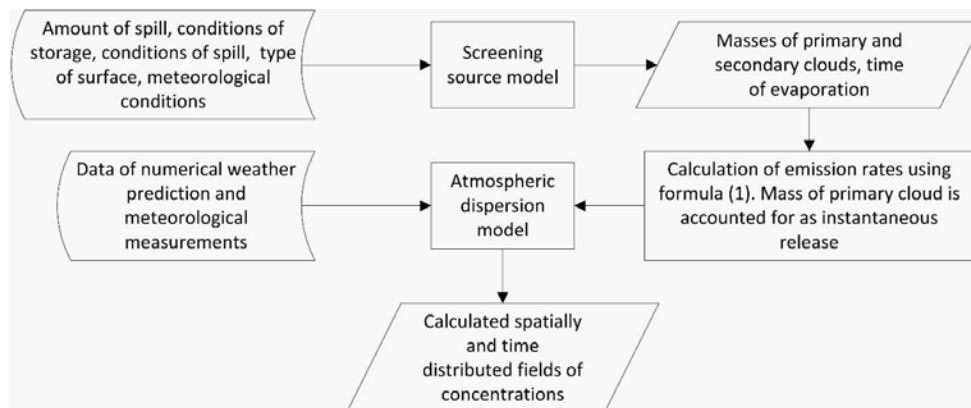


Fig. 1. Flowchart of calculation of air pollution based on the combined application of screening models to assess the characteristics of sources of emissions of hazardous chemicals with state-of-art models of atmospheric transport

## MODEL VERIFICATION

Let us compare the results of modeling the depth of propagation of a chlorine cloud after a rapid spill of 67 tons of this substance, performed by the “Povitrya”

system, with calculations of international atmospheric transport models [17], specially designed to calculate the propagation of heavy gases, such as chlorine. The meteorological conditions of the modeling scenario [17] were as follows: wind speed of 3 m/s, stability category ‘F’ (very stable atmosphere), and air temperature of 25 °C. When modeling such an event by the “Povitrya” system according to the method [3], the depth of the cloud propagation zone is 20 km, and the evaporation time is 90 min.

The range of concentrations calculated by eight models [17] at different distances from the source is shown in Fig. 2. As can be seen from the presented data, the range of concentrations calculated by the models in [17] for a distance of 25 km is from 2 to 30 ppm. According to the American Industrial Hygiene Association (AIHA), this range includes the maximum concentration that does not cause irreversible health effects when exposed for one hour ERPG2=3 ppm, and the maximum concentration of chlorine that does not threaten life ERPG3=20 ppm, [18]. Therefore, according to the calculations of the models in [17], at a distance of 25 km, it is necessary to take countermeasures for the event under consideration, which is consistent with the cloud propagation depth of 20 km according to the results of the “Povitrya” system.

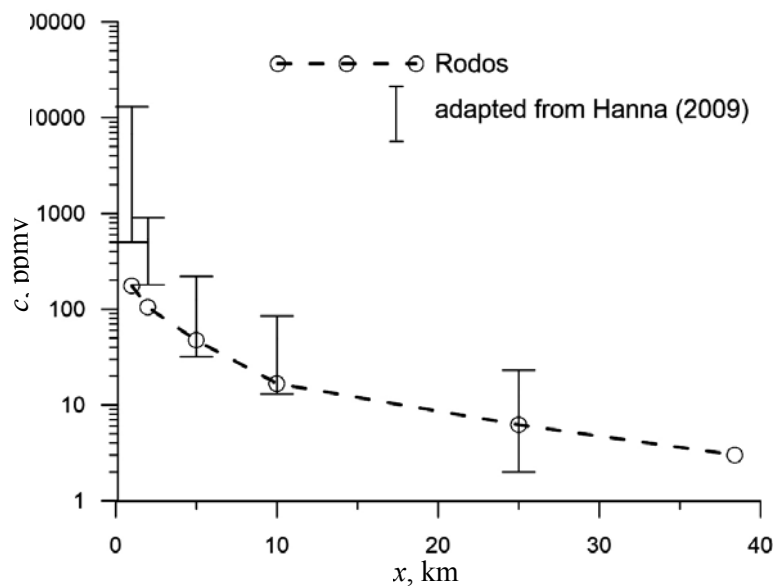


Fig. 2. Maximum concentrations of chlorine in the air calculated by the RODOS system on the cloud symmetry axis for the conditions of the computational experiment according to [17] and the duration of the source calculated by the “Povitrya” system; concentration ranges according to the calculations of 8 models in [17] are shown by vertical lines

Fig. 2 also shows the results of calculating the maximum chlorine concentrations by the RODOS system at the same distances as the models in [17]. The emission source in the RODOS system was set according to the formula (1), where  $\tau=90$  min according to the calculation of the “Povitrya” system, and the mass of the primary cloud was neglected. The emission height was assumed to be 10 m since the emission height in the RODOS system cannot be less than 10 m. As can be seen from the data presented in the figure, the concentrations calculated by RODOS are underestimated compared to other models according to [17] and are outside the ranges concentrations of these models up to a distance of 5 km.

This difference is most likely due to buoyancy effects that affect meteorological fields in the vicinity of the cloud when the density of the cloud is significantly different from the density of the surrounding air [19]. These effects may be important for modeling the propagation of chlorine at high concentrations since the latter is a heavy gas whose density at atmospheric pressure is about 2.2 times that of air. Buoyancy effects are taken into account by models [17], which are specially designed for calculating heavy gases and are ignored by both the RODOS system and almost all ATMs used to calculate atmospheric transport at a distance of more than 30 km since taking into account buoyancy effects requires separate approaches to modeling.

Let us evaluate the significance of buoyancy effects for this problem. In [19], a parameter was introduced that characterizes the significance of such effects depending on the emission conditions and meteorological conditions. According to [19], buoyancy effects can be neglected if the following condition is satisfied:

$$\Pi = \frac{(\dot{g}'_0 q_0 / D)^{1/3}}{U} < 0.15. \quad (2)$$

Here  $U$  is the wind speed,  $D$  is the source diameter,  $q_0$  is the volumetric gas flow rates in the source ( $\text{m}^3/\text{s}$ ),  $\dot{g}'_0 = g(\rho_g - \rho_a)/\rho_a$  is buoyancy,  $g$  is the gravitational acceleration,  $\rho_g$  is the gas density at atmospheric pressure,  $\rho_a$  is the air density. To use formula (2), we estimate the parameters:  $\dot{g}'_0 = 12 \text{ m/s}^2$ ,  $q_0 \approx 4.5 \text{ m}^3/\text{s}$  (the estimate of the evaporation time according to the data of the "Povitrya" system was used),  $U = 3 \text{ m/s}$ . To estimate the source diameter, we use the approach [3], in which it is assumed that during a spill on the real surface of the Earth, the spill diameter ceases to increase at a spill basin height of less than 0.05 m due to the impact of minor surface irregularities. Using this estimate, it is possible to obtain an estimate of the spill diameter  $D = 33 \text{ m}$ . Substituting these data into formula (2), we obtain  $\Pi = 0.4$ , so the effects of buoyancy are important for this problem. When estimating pollution levels near the source the buoyancy effects could be properly accounted for by using specialized models of the heavy gas dispersion, such as those used in [17].

As can be seen from Fig. 2, at distances greater than 5 km, the results calculated by RODOS fall within the range of model calculations [17]. If at a distance of 5 km the results of RODOS are at the lower limit of the calculation range of other models, then at a distance of 25 km they are already inside the range. The explanation for this is that the effects of buoyancy decrease with distance due to the mixing of the cloud with the surrounding air [19]. As a result, the distance to a dangerous concentration of ERPG2=3 ppm according to the calculations of the RODOS system is 38 km, which is within a significant interval of the corresponding distances according to the models [17]. The lower limit of this interval is  $\approx 20 \text{ km}$  (obtained by interpolating the lower limits of the intervals depicted in Fig. 2 to a value of 3 ppm), and the upper limit is  $\approx 46 \text{ km}$  (obtained by extrapolating the upper limits of the intervals in Fig. 2 to a value of 3 ppm).

As can be seen from Fig. 2, the difference between the calculations of all models is large — almost 10 times, and the level of discrepancy practically does not decrease with distance, although the effects of buoyancy cease to affect cloud

propagation. The corresponding discrepancies between the distances from the source to the places where the fixed concentration is reached are also very significant, as mentioned above. In addition to negative cloud buoyancy effects and differences in the results of evaporation submodels, such differences in calculated concentrations could be explained by the difficulties in accurately calculating the propagation of pollutants near the surface, where horizontal velocity gradients are large. This leads to the effect of “shear dispersion” and the corresponding additional dilution of the substance concentration [20].

## RESULTS OF CALCULATIONS OF THE AMMONIA EMISSION IN CHERNIHIV

By using the proposed methodology, we carried out modeling of the spread of ammonia vapors after damage to the ammonia pipeline during the shelling of one of the industrial facilities in Chernihiv on March 23, 2002 [21]. Since the total volume of ammonia emission is unknown, in this work the estimate  $m_0=50$  tons was accepted. Such a mass corresponds to a spill through a hole with a diameter of 10 cm at a pressure in the container of 100 psig = 689 kPa for about 4 min [17]. The pressure of 689 kPa is close to the typical pressure in ammonia pipelines of 862 kPa [22]; therefore, the error in the given estimate is determined mainly by the unknown spill time.

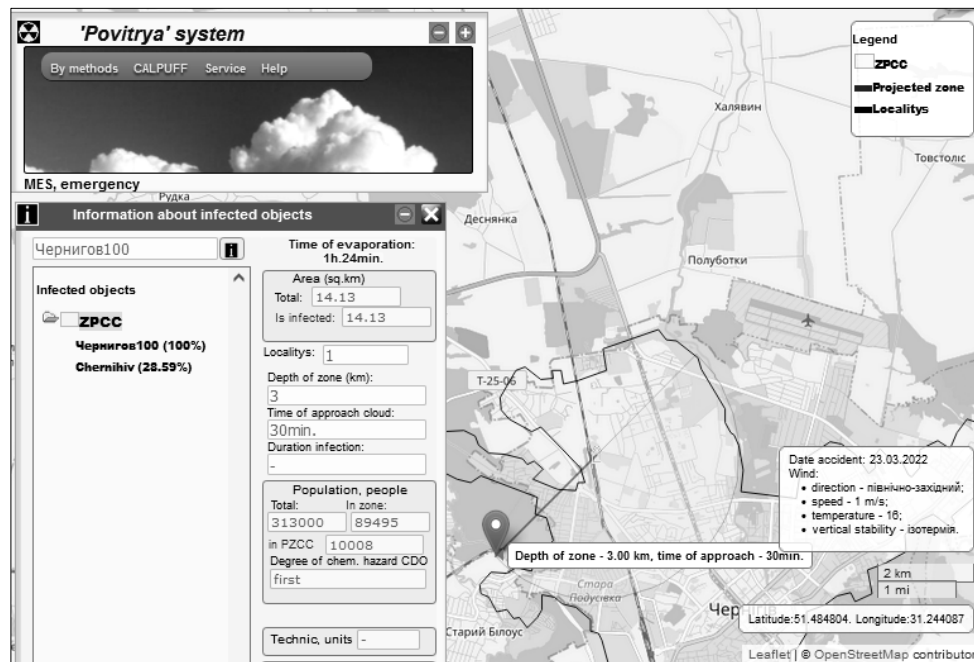


Fig. 3. The results of the calculation by the “Povitrya” system of the consequences of an accident in Chernihiv with a leak of 50 tons of ammonia

Fig. 3 shows the results of calculations of the “Povitrya” system for the scenario under consideration. The meteorological conditions in the calculations were identical to those observed on the day and time of the accident: northwest wind, wind speed 1 m/s, air temperature 16 °C, stability category C. The calculated cloud propagation depth was 3 km. The calculated evaporation time according to

the model [3] was  $\tau = 84$  min, and the corresponding value obtained according to the model [4]  $\tau = 62$  min. According to the calculations, the mass of the primary cloud was  $m_1 = 11.7$  tons, that is, approximately 23% of the total emission volume.

To simulate the propagation of the cloud by the RODOS system, the evaporation time of the cloud was taken to be  $\tau = 60$  min. The simulation was carried out on the basis of meteorological data calculated by the WRF–Ukraine weather forecasting system. The RODOS system performed calculations on two computational grids:

1) calculation with the minimum resolution of the spatial grid, which is equal to 50 m, and the total size of the calculation area is  $40 \times 40$  km (hereinafter referred to as “grid 1”);

2) calculation with a coarser minimum resolution of the spatial grid, which is equal to 500 m, and the total size of the computational area is  $400 \times 400$  km (hereinafter referred to as “grid 2”).

Fig. 4 shows the calculated concentration field one hour after the start of the emission when the concentration values in the near zone around the emission are maximum (this is the point at which the emission stops). According to [18], the maximum concentration of ammonia that does not threaten the occurrence of irreversible or other serious health effects is  $ERPG2 = 150$  ppm =  $105$  mg/m<sup>3</sup>. The isoline of the corresponding concentration is shown in Fig. 3. The maximum distance from the emission point to this isoline, according to RODOS calculations, was:  $L = 1.8$  km. The corresponding value of the cloud propagation depth calculated by the “Povitrya” system according to the model [3] is 3 km. This discrepancy can be considered small given the typical discrepancies between models discussed in the previous section.

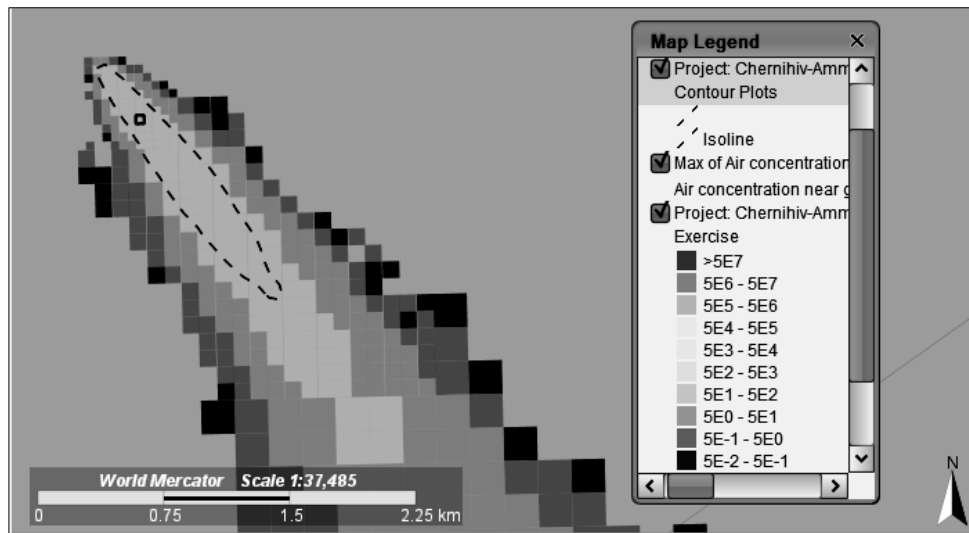


Fig. 4. Ground air pollutant concentrations calculated by the RODOS system (mg/m<sup>3</sup>) as a result of the emission of 50 tons of ammonia one hour after the start of the emission; the calculation was performed on grid 1 for meteorological conditions on March 23, 22; dashed isoline corresponds to the concentration of 150 ppm =  $105$  mg/m<sup>3</sup> (normative value of ERPG2 according to [18])

Fig. 4 shows a cell with the maximum concentration value, which is equal to  $910$  mg/m<sup>3</sup> and is reached at a distance of 200 m from the source. According to [18],

ammonia concentrations of  $1050 \text{ mg/m}^3$  and higher are deadly. In the considered scenario, such concentrations are not achieved, which is consistent with the above media report that, according to preliminary data, there were no casualties caused by this event. But it should be noted that, according to the information [3], a conservative estimate when calculating an accident at an ammonia pipeline or a product pipeline can be considered an outflow of a substance in the amount of 500 tons. When calculating such a conservative scenario by the RODOS system for meteorological conditions on March 23, 2022, the size of the affected area in the direction of the wind, which corresponds to the above value of the lethal concentration, is 1.8 km.

Fig. 5 shows the isolines of the concentration of  $0.2 \text{ mg/m}^3$  calculated by the RODOS system on grid 2 for different time moments after the accident. According to the Order of the Ministry of Health of Ukraine No. 52 dated December 14, 2020, the value of  $0.2 \text{ mg/m}^3$  for ammonia is the maximum permissible concentration (MPC). As can be seen from the location of the isolines shown in Fig. 5, at first the cloud moves to the southeast, but over time the wind changes direction, and the cloud continues to move in a northeasterly direction. MPC exceedances are observed during 9.5 hours of cloud movement at distances up to 75 km, after which concentrations become lower than MPC. Owing to the dilution of the cloud

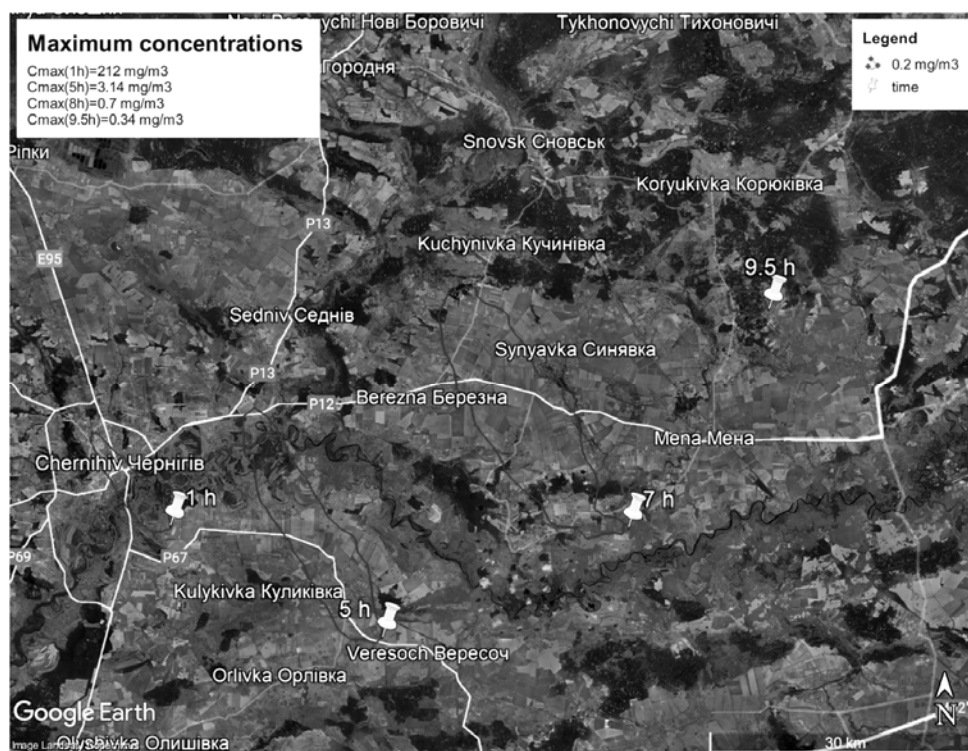


Fig. 5. Isolines of surface concentrations of ammonia corresponding to the standard  $\text{MPC}=0.2 \text{ mg/m}^3$ , calculated by the RODOS system on grid 2 for different time moments after the accident in Chernihiv (1, 5, 7, 9.5 h) for meteorological conditions on March 23, 2022. For each time, the figure shows the maximum concentrations in the cloud

and the corresponding increase in its horizontal and vertical dimensions, the maximum concentrations in the cloud rapidly decrease. The maximum value achieved within each isoline depicted in Fig. 5 is also shown in the figure. Calculations



lation on a coarse grid leads to a strong underestimation of the concentration near the source: one hour after the start of the emission, the maximum value on grid 2 is 212 mg/m<sup>3</sup>, while on grid 1 the corresponding value is 910 mg/m<sup>3</sup>. However, over time, this difference disappears. For example, 5 hours after the start of the emission, the maximum concentrations calculated on both grids differ by less than 10%.

The time dependence of the maximum concentrations calculated by the RODOS system on grids 1 and 2 with an interval of 1 hour is shown in Fig. 6. Since the results are presented only for the times when the cloud is in the computational domain, the time series corresponding to the plot for grid 2 is longer than for grid 1. To test the sensitivity of the results to the model used, in addition to the LASAT model that was predominantly used in this work. Fig. 6 shows the results of the DIPCOT model, also available in the RODOS system. Although DIPCOT also uses a Lagrangian approach to modeling stochastic particle motions in a turbulent medium, in contrast to LASAT, DIPCOT particles are characterized by finite sizes, that is, they are “puffs”, as discussed above.

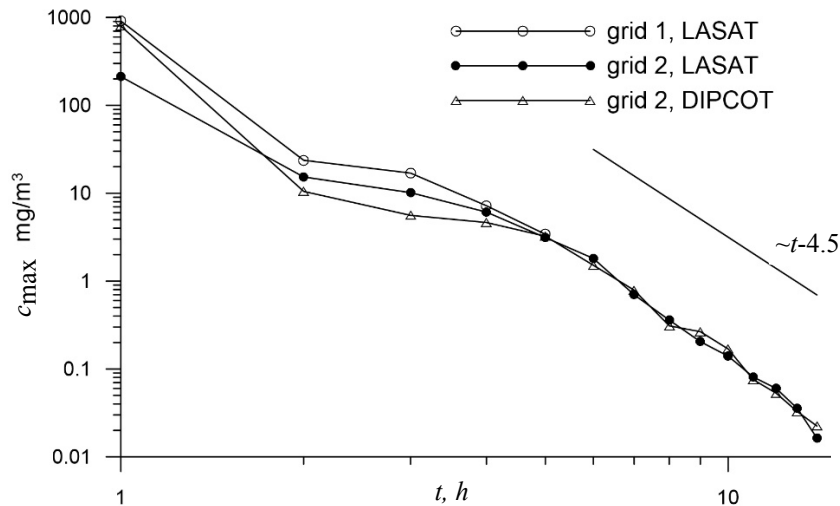


Fig. 6. Dependence of the highest concentrations in the computational area on time after the start of the emission according to the results of RODOS calculations on grids 1 and 2; for grid 2, the results of two RODOS models are presented: LASAT and DIPCOT; the results are presented only for the times when the cloud is inside the computational domain

From the graphs presented, it can be seen that in all calculations three time intervals can be distinguished: 1)  $1 < t \leq 2$  hours, the maximum concentration decreases rapidly; 2)  $2 < t \leq 4h$ , the concentration decreases slowly; 3)  $t > 4h$ , the concentration drops rapidly again. The interval  $0 < t \leq 1 h$  is not considered, since during this period an emission occurs and the maximum concentration is approximately constant. As can be seen from the above results, during intervals 1) and 3) the maximum concentration decreases with time approximately as:

$$c_{\max} \sim t^{-4.5}, \quad (3)$$

and the corresponding asymptotics is also shown on the graph. This dependence of the maximum concentration on time is in accordance with the classical results of the theory of turbulent motion of particles, according to which the vertical and

horizontal dimensions of a medium particle after an instantaneous emission increase following the dependence [23]:

$$\sigma \sim t^{3/2}. \quad (4)$$

Indeed, in the case of an instantaneous emission of a substance, if the height of the emission (puff) above the Earth's surface is much less than its characteristic vertical size  $\sigma_v$ , that is typical for the long-term existence of the puff [24], the maximum concentration on the Earth's surface will be determined by the dependence  $c_{\max} \sim \sigma_h^{-2} \sigma_v^{-1}$ . Substituting expression (4) into the last relation, we obtain the asymptotic dependence (3), which agrees with the graphs in Fig. 5 at time intervals 1) and 3). Some deviations from this dependence on time intervals 1) and 3) are explained by the fact that for turbulent dispersion in real meteorological fields that vary in time and space, no analytical relationships for the sizes of puffs with time can be fulfilled exactly.

It should be noted that, according to theoretical studies in [23], when the duration of the movement of puffs is much longer than the Lagrangian integral time scale of turbulence in the atmosphere (i.e., several hours after the emission), the law of growth of puff's sizes (4) should be replaced by more slow dependence  $\sigma \sim t^{3/4}$ . Then the time dependence of the maximum concentration would be described by the relation  $c_{\max} \sim t^{-2.25}$ , which obviously does not occur according to the calculations of both RODOS models, the data of which are presented in Fig. 6. Such a discrepancy between theory and model calculations may be the subject of further research.

A sharp slowdown in the rate of fall of the maximum concentration occurs in time interval 2) from the second to the fourth hour after the emission, exactly when the direction of the cloud movement changes from southeast to northeast. This is associated with the advancement of the pressure trough from the north and the corresponding change in wind direction. At this time interval, the wind speed drops below 0.5 m/s for a short time. Hence calm conditions arise when the existing cloud moves almost without mixing. Within two hours, a stable transport is established in the northeast direction, the wind speed increases to 3.5 m/s, and mixing is restored.

## CONCLUSIONS

In the present work, we studied the possibility of the combined application of screening models for assessing the characteristics of emission sources during accidents at hazardous chemicals storage and transportation facilities with complex models of atmospheric transport, functioning as part of state-of-art environmental decision support systems, for calculating air pollution in a wide range of spatial and time scales – from ~100 m to ~100 km. The evaporation time and the mass of the primary cloud in case of an accidental spill of a substance are calculated by using the web system "Povitrya" based on the calculation of screening models for assessing affected areas in emergency situations at industrial facilities. These parameters are then used to set the emission intensity and calculate the atmospheric transport by the RODOS system of the European Union for nuclear emergency response. The advantage of using simplified screening models in comparison with

the more complex models of releases results from the fact that relatively few parameters need to be specified for the application of simplified models (volume of the spill, storage and meteorological conditions, type of terrain). The values of the corresponding parameters could be estimated in real-time.

The chlorine concentrations calculated in this way for the scenario of an accidental release are compared with the results of international models specially designed to calculate the spread of heavy gases. It was found that the concentrations calculated by the RODOS system for this scenario are consistent with the results of other models at distances of more than 5 km and are underestimated at shorter distances. This is due to the effect of negative buoyancy on the propagation of the chlorine cloud. With distance, the effects of buoyancy cease to affect cloud propagation, and as a result, the distance to a dangerous concentration, according to RODOS calculations, is within the interval of the results of other models.

The proposed approach was used to analyze the spread of ammonia vapor as a result of damage to the pipeline in Chernihiv on March 23, 2022, caused by hostilities. The amount of ammonia emission for this event is unknown. A conservative estimate of the emission of 500 tons for this event leads to an estimate of the length of the affected area according to the RODOS system up to 1.8 km from the source in the direction of the wind. Based on information from the media about the absence of victims, an assumption was made about the emission of 50 tons of ammonia. Based on the calculations of the RODOS system using the data of the WRF–Ukraine meteorological forecasting system, it was estimated that, due to the spread of the cloud, the maximum permissible concentration of ammonia, MPC = 0.2 mg/m<sup>3</sup>, was exceeded at distances up to 75 km from the point of emission. Exceeding the MPC in a certain area means that the cloud could have a dangerous effect on the organisms located in the corresponding area. An assessment of the degree of such exposure can be made based on the results of the models used in this work, but requires the involvement of experts in toxicology and is not the subject of this article.

The temporal behavior of the highest concentrations calculated by the RODOS system was analyzed. The dependence of maximum concentrations on time is asymptotic, close to  $c_{\max} \sim t^{-4.5}$ , which is consistent with the classical asymptotic relation  $\sigma \sim t^{3/2}$  for the time dependence of the size of puffs (instant emissions) during their turbulent mixing in the atmosphere on time intervals smaller than the Lagrangian integral time scale of turbulence  $t_L$ . However, since the propagation was calculated for time intervals up to 15 hours, which is more than  $t_L$ , in future theoretical studies on atmospheric dispersion, it would be necessary to clarify the possibility of fulfilling the dependence  $\sigma \sim t^{3/2}$  on large time intervals.

The method of combining screening models for evaluating the characteristics of sources in the case of severe man-made accidents at industrial facilities with state-of-art models of atmospheric transport studied in this work can be used to forecast the consequences of such events for the environment and people, for the development of new and improved decision support systems for environmental safety management, and to assess risks due to operation of industrial objects.

## REFERENCES

1. J.E. Cudney-Black et al., *Hazardous Chemical Inventory Guidelines, Purpose and Process*. Sandia National Lab., Albuquerque, New Mexico, USA, Rep. SAND2021-8639, 2021. doi: 10.2172/1821974.
2. C.J.H. van den Bosch and R.A.P.M. Weterings, Eds., *Methods for the calculation of physical effects due to releases of hazardous materials (liquids and gases) – ‘Yellow Book’* (Publication series on dangerous substances); 3d ed. The Hague, The Netherlands: Director-General of Labor : Committee for the Prevention of Disasters, 2005.
3. “Recommendations on forecasting the consequences of spills (emissions) of hazardous chemicals at industrial facilities and transport,” (in Ukrainian), *Ministry of Ukraine of Emergencies and Affairs of Population Protection from the Consequences of Chernobyl catastrophe*. Kyiv, 2001. Accessed on: Jun 6, 2022. [online]. Available: <https://zakon.rada.gov.ua/laws/show/z0326-01>.
4. “Method of assessment of chemical situation following accidents at facilities containing toxic substances,” (in Russian), *Ministry of Defense of USSR*. M., 1989.
5. U. Janicke and L. Janicke, “Lagrangian particle modeling for regulatory purposes; A survey of recent developments in Germany,” in *Proc. of the 11th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, HARMO 2007, Cambridge, UK, 2007*, pp. 109–113.
6. S. Andronopoulos, E. Davakis, J.G. Bartzis, and I. Kovalets, “RODOS meteorological pre-processor and atmospheric dispersion model DIPCOT: a model suite for radionuclides dispersion in complex terrain,” *Radioprotection*, vol. 45, no. 5, pp. 77–84, 2010. doi: 10.1051/radiopro/2010017.
7. C. Landman, J. Pasler-Sauer, and W. Raskob, “The decision support system RODOS,” in *The Risks of Nuclear Energy Technology. Science Policy Reports*. Heidelberg, Germany: Springer-Berlin, 2014, pp. 337–348. doi: 10.1007/978-3-642-55116-1\_21.
8. C. Gao et al., “Two-way coupled meteorology and air quality models in Asia: a systematic review and meta-analysis of impacts of aerosol feedbacks on meteorology and air quality,” *Atmospheric Chemistry and Physics*, vol. 22, pp. 5265–5329, 2002. doi: 10.5194/acp-22-5265-2022.
9. S. Khan and H. Quamrul, “Review of developments in air quality modelling and air quality dispersion models,” *J. of Environmental Engineering and Science*, vol. 16, no. 1, pp. 1–10, 2021. doi: 10.1680/jenes.20.00004.
10. I. Kovalets and O. Romanenko, “Use of Nuclear Emergency Response System for Assessment of Transboundary Transfer and Radiological Risks of the Potential Accidental Releases at Khmelnytsky NPP,” *Advances in Intelligent Systems and Computing*, vol. 1265, pp. 3–12, 2021. doi: 10.1007/978-3-030-58124-4\_1.
11. Y. Kyrylenko et al., “Source Term Modelling for Event with Liquid Radioactive Materials Spill,” *Studies in Systems, Decision and Control*, vol. 298, pp. 261–279, 2020. doi: 10.1007/978-3-030-48583-2\_17.
12. J.G. Powers et al., “The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions,” *Bulletin of the American Meteorological Society*, vol. 98, no. 8, pp. 1717–1737, 2017. doi: 10.1175/BAMS-D-15-00308.1.
13. A. Khalchenkov and I. Kovalets, “Evaluation of Spectral/Grid Nudging Methods for Weather Analysis and Forecasting in Kyiv Region with the Use of WRF Mesoscale Meteorological Model,” *Advances in Intelligent Systems and Computing*, vol. 1265, pp. 13–23, 2021. doi: 10.1007/978-3-030-58124-4\_2.
14. I.V. Kovalets et al., “Povitrya web-based software system for operational forecasting of atmospheric pollution after manmade accidents in Ukraine,” *Science and Innovation*, vol. 13, no. 6, pp. 11–22, doi: 10.15407/scin13.06.013.

15. System "Povitrya". [Online]. Available: [https://cloud-2.bitp.kiev.ua/airsystem\\_english/airsystem.html](https://cloud-2.bitp.kiev.ua/airsystem_english/airsystem.html).
16. M. Oksanych, V.P. Bespalov, and I.V. Kovalets, "Comparison of methods for calculating the parameters of evaporation of hazardous liquids from the surface of the spill basin on a hard surface," (in Ukrainian), *Mathematical Machines and Systems*, no. 2, p. 49–59, 2017.
17. S.R. Hanna, O.R. Hansen, M. Ichard, and D. Strimaitis, "CFD model simulation of dispersion from chlorine railcar releases in industrial and urban areas," *Atmospheric Environment*, vol. 43, no. 2, pp. 262–270, 2009. doi: 10.1016/j.atmosenv.2008.09.081.
18. *OSHA Occupational Chemical Database*, US Department of Labor, Occupational Safety and Health Administration (OSHA), May 2022. [Online]. Available: <https://www.osha.gov/chemicaldata>.
19. R.E. Britter, "Atmospheric dispersion of dense gases," *Annual Reviews of Fluid Mechanics*, vol. 21, pp. 317–344, 1989. doi: 10.1146/annurev.fl.21.010189.001533.
20. A.H. Kumar, S.J. Thomson, T.R. Powers, and D.M. Harris, "Taylor dispersion of elongated rods," *Physics Review Fluids*, vol. 6, no. 9, Sept. 2021, Art. no. 094501. doi: 10.1103/PhysRevFluids.6.094501.
21. "Aggressors' shell damaged an ammonia pipeline in Chernihiv," *Ukrayinska Pravda*, 23.03.2022. [Online]. Available: <https://www.pravda.com.ua/eng/news/2022/03/23/7334017/>.
22. S. Papavinasam, "Chapter 2 - Oil and Gas Industry Network," in *Corrosion Control in the Oil and Gas Industry*, S. Papavinasam, Ed., Amsterdam, The Netherlands, Gulf Professional Publishing, 2014, ch. 2, pp. 41–131, doi: 10.1016/B978-0-12-397022-0.00002-9.
23. T. Mikkelsen, S. Thykier-Nielsen, and S. Hoe, "Chapter 2.16 Medium-range puff growth," in *Air Pollution Modeling and Its Application XVIII* (Developments in Environmental Science, vol. 6), C. Borrego and E. Renner, Eds., Amsterdam, The Netherlands, Elsevier, 2007, pp. 243–252, doi: 10.1016/S1474-8177(07)06216-X.
24. T. Mikkelsen, "Atmospheric Dispersion: Basic," in *Encyclopedia of Environmetrics Vol.1*, A.H. El Shaarawi, W.W. Piegorsch, Eds., Chichester, UK, John Wiley & Sons, 2002, pp. 85–106, doi: 10.1002/9780470057339.vaa025m.

Received 06.06.2022

#### INFORMATION ON THE ARTICLE

**Ivan V. Kovalets**, ORCID: 0000-0002-5553-3299, The Institute of Mathematical Machines and Systems Problems of the Ukraine National Academy of Science, Ukraine, e-mail: [ivkov084@gmail.com](mailto:ivkov084@gmail.com)

**Viacheslav P. Bespalov**, ORCID: 0000-0001-6841-1533, The Institute of Mathematical Machines and Systems Problems of the Ukraine National Academy of Science, Ukraine, e-mail: [bespalovviacheslav@gmail.com](mailto:bespalovviacheslav@gmail.com)

**Svitlana Ya. Maistrenko**, ORCID: 0000-0002-1316-9732, The Institute of Mathematical Machines and Systems Problems of the Ukraine National Academy of Science, Ukraine, e-mail: [maistrsv@ukr.net](mailto:maistrsv@ukr.net)

**Oleg I. Udovenko**, ORCID: 0000-0001-5679-8291, Ukrainian Centre for Water and Ecology Issues, Ukraine, e-mail: [oleg.udovenko@gmail.com](mailto:oleg.udovenko@gmail.com)

**ВИКОРИСТАННЯ СИСТЕМ ПІДТРИМАННЯ РІШЕНЬ З ЕКОЛОГІЧНОЇ БЕЗПЕКИ ДЛЯ МОДЕЛЮВАННЯ ЗАБРУДНЕННЯ АТМОСФЕРИ ВНАСЛІДОК ХІМІЧНИХ АВАРІЙ / І.В. Ковалець, В.П. Беспалов, С.Я. Майстренко, О.І. Удовенко**

**Анотація.** Досліджено можливість комбінованого застосування скринінгових моделей для оцінювання характеристик джерел у разі аварій на об'єктах зберігання небезпечних речовин зі складними моделями атмосферного перенесення у складі сучасних систем підтримання рішень для розрахунку атмосферного забруднення у широкому діапазоні просторових і часових масштабів. Час випаровування у разі аварійного розливу, оцінений скринінговими моделями, використовується для задання інтенсивності емісії і розрахунку атмосферного поширення системою ядерного аварійного реагування RODOS. Для аварії в Чернігові 23.03.2022 оцінено, що перевищення граничної допустимої концентрації аміаку  $0.2 \text{ мг/м}^3$  відбувається на відстанях до 75 км від джерела. Залежність розрахованих максимальних концентрацій від часу має асимптотичний характер близький до  $c_{\max} \sim t^{-4.5}$  до 15 год після викиду, що узгоджується з асимптотичним співвідношенням  $\sigma \sim t^{3/2}$  для часової залежності розмірів хмар у разі турбулентної дисперсії миттєвих викидів.

**Ключові слова:** атмосферна дисперсія, система RODOS, небезпечні речовини, аміак, LASAT, DIPCOT, система «Повітря».

## ПРОЦЕС КЕРУВАННЯ ЗАХИЩЕНІСТЮ ДАНИХ ПІД ЧАС ВІДДАЛЕНОЇ БІОМЕТРИЧНОЇ АВТЕНТИФІКАЦІЇ

А.А. АСТРАХАНЦЕВ, Г.Є. ЛЯШЕНКО

**Анотація.** Системи віддаленої біометричної автентифікації за останній час набули значного поширення через необхідність користування загальними пристроями та виконання платежів через Інтернет. Оскільки саме біометричні методи більш зручні для користувачів і нині швидко замінюють паролі, то стає актуальним завданням передавання біометричної інформації відкритою мережею без її компрометації. Метою роботи є модернізація системи віддаленої автентифікації для підвищення рівня прихованості і захищеності біометричних даних користувача. Запропоновано застосування найкращих за критерієм захищеності методів формування біометричного шаблону, методів мережевої стеганографії для підвищення прихованості та впровадження інтелектуальної системи прийняття рішень. Такі вдосконалення дозволять підвищити захищеність і прихованість даних для процесу віддаленої автентифікації.

**Ключові слова:** біометричний шаблон, віддалена автентифікація, атаки, мережева стеганографія.

### ВСТУП

Натепер дедалі більшого попиту набули інтернет-магазини, онлайн-банкінг та інші послуги, під час яких користувач користується віддаленим передаванням своєї особистої інформації. Під час онлайн-оплати послуг (магазини, комунальні платежі, купівля квитків та ін.) дуже важливим є захист персональних даних користувача. Дедалі частіше під час віддаленої автентифікації користувача використовується біометрична автентифікація [1, 2]. Вона базується на використанні таких невід’ємних та унікальних для кожної людини фізичних характеристик, як відбитки пальців, зображення райдужної оболонки ока, геометрії обличчя, або поведінкових характеристик. Оскільки ці характеристики не можуть бути відновлені або замінені у випадку викрадення чи втрати, то завдання забезпечення захищеності даних під час передавання мережею або в процесі одностороннього перетворення даних для унеможливлення використання зловмисником є дуже важливими.

**Мета роботи:** удосконалення процесів, що відбуваються під час віддаленої біометричної автентифікації та під час передавання автентифікаційних даних мережею, огляд можливих атак у мережі та обрання методів, які допоможуть підвищити захищеність даних під час віддаленої автентифікації для передавання відкритими каналами зв’язку.

### БІОМЕТРИЧНА АВТЕНТИФІКАЦІЯ: ЗАГАЛЬНІ ПРИНЦИПИ ТА МОЖЛИВІ АТАКИ

У загальному вигляді схему біометричної автентифікації зображено на рис. 1, а, б. Система автентифікації в базі даних зберігає біометричні шаб-

лони зареєстрованих користувачів, дані про них та інструкції щодо режиму доступу певних об'єктів. Біометричний датчик на вході зчитує унікальні біометричні характеристики користувача, система порівнює їх з тими, що внесені до бази даних, авторизує користувача і в разі збігу характеристик користувача з шаблоном, що внесений до бази даних, надає рішення щодо допуску до певної інформації/об'єктів/тощо.

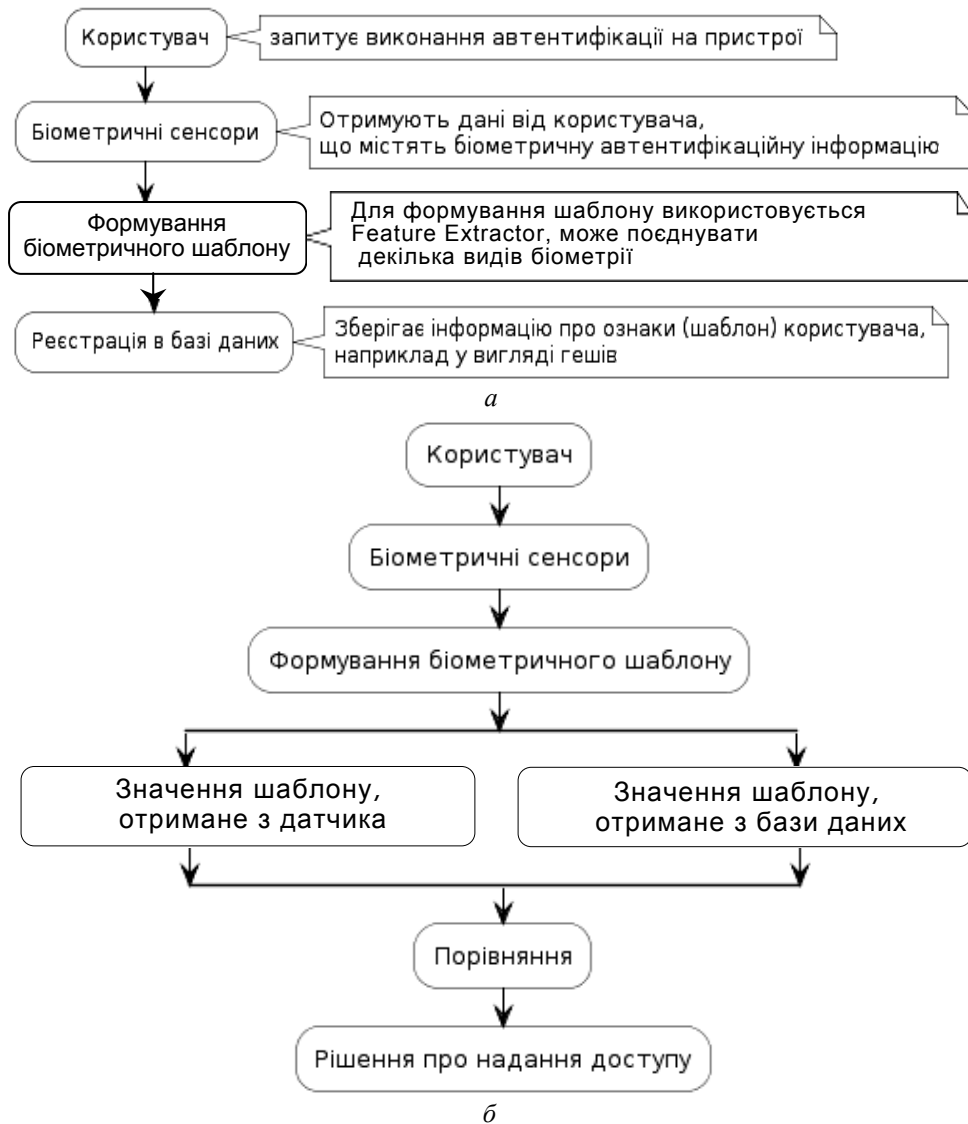


Рис. 1. Узагальнена схема біометричної автентифікації: фаза реєстрації (а), фаза перевірки (б)

Під час віддаленої автентифікації дані, що мають автентифікувати користувача, передаються мережею, під час чого можуть бути спотворені завадами в каналах зв'язку та скомпрометовані в результаті різних типів атак. Біометричні характеристики людини є унікальними і невід'ємними. Це дає великі переваги для правильного надання доступу з використанням цих характеристик, але вони не можуть бути замінені, тому передавати їх у відкритому вигляді або навіть зашифрованому неможливо.



На систему віддаленої автентифікації можливі різні типи атак (рис. 2). На самому початку системи передавання даних можлива фальсифікація даних (spoofing attack) — використання фальсифікованих біометричних характеристик користувача, поновлення та використання старих даних, які використовувалися раніше під час автентифікації. Також існує ймовірність атаки у вигляді несанкціонованого доступу до сформованого біометричного шаблону під час автентифікації, його підміна або підміна шаблону (substitution attack) [1, 13], який зберігається в базі даних. Небезпечною є атака маскаррад (masquerade attack), коли цифровий образ може бути створений із шаблону біометричного образу. Існує ймовірність впливу з метою підміни рішення під час порівняння біометричних шаблонів.



Рис. 2. Можливі вразливості в системі віддаленої біометричної автентифікації

Окрім зазначених атак, під час передавання даних каналом зв'язку є загроза того, що дані будуть перехоплені.

Під час спуфінгових атак зловмисник для отримання доступу в систему використовує штучно створені біометричні ознаки. Наприклад, маску обличчя, надруковане зображення райдужної оболонки ока тощо, або імітує поведінкові характеристики, що використовуються під час динамічної автентифікації, наприклад, динамічний підпис. Такі загрози виникають на етапі роботи з датчиками, що розпізнають особу, тому для захисту від таких загроз потрібно вжити заходів, що будуть попереджати від розпізнавання підроблених зразків. Прикладом біометричних властивостей, які мають високий рівень захищеності від підроблень, є такі властивості, як температура, електрична провідність, пульсоксиметрія та опір шкіри.

Для боротьби зі спуфінгом на рівні пристроїв використовують апаратні методи. Вони полягають в інтеграції у сканер спеціальних апаратних пристроїв, які дозволяють розпізнавати конкретні характеристики живих біомет-

ричних зразків (рухи ока, тепло пальців та ін.); вони також можуть перевіряти реакцію на зовнішні сигнали, що потребують наявності певного користувача. Наприклад, для того, щоб попередити використання підробних відбитків пальців, що створені зі штучного матеріалу сканери відбитків пальців, застосовують механізми виявлення підробних відбитків.

Для вирішення проблеми використання підроблених зображень використовують ідентифікацію користувачів за рухами зіниць, які виникають у разі змін їх розмірів.

На рівні функцій використовують програмні методи. Ці методи інтегруються в систему після сканерів. Робота таких методів полягає в екстракції ознак і вони працюють уже з послідовностями ознак, отриманих за певний проміжок часу.

Під час перших етапів роботи системи автентифікації важливо уникнути можливості відновлення старих, уведених іншим користувачем даних для отримання несанкціонованого доступу.

Для здійснення біометричної автентифікації після сканування певних біометричних ознак формується біометричний шаблон. Біометричний шаблон — це цифрове подання даних, вилучених з біометричного зразка. Вони зберігаються в базі даних і використовуються під час порівняння для автентифікації. Існує ризик поновлення створеного шаблону та використання старих даних. Є ймовірність підміни шаблону на інший та перехоплення шаблону в каналі зв'язку у процесі передавання, тому важливим завданням є захист шаблонів.

## **ОГЛЯД ВИМОГ ТА ПІДХОДІВ ДО ЗАХИСТУ БІОМЕТРИЧНОГО ШАБЛОНУ**

Ідеальна біометрична схема захисту шаблону повинна мати такі чотири властивості [7].

1. Різноманітність: безпечний шаблон не повинен дозволяти порівняльний пошук по базах даних, тим самим забезпечуючи конфіденційність користувача.

2. Можливість ануляції: має бути простою для відкликання скомпрометованого шаблону та перевипуску нового, заснованого на тих же біометричних даних.

3. Безпека: отримання оригінальної біометричної інформації із сформованого шаблону повинно бути обчислювально важким. Ця властивість перешкоджає відновленню біометричних ознак з викраденого шаблону.

4. Продуктивність: схема захисту біометричного шаблону не повинна погіршити продуктивність розпізнавання.

Основним викликом розроблення біометричної схеми захисту шаблону, який задовольняє всі згадані вимоги, є необхідність оброблення мінливих даних користувача.

Нагадаємо, що декілька зображень однієї біометричної ознаки не призводять до того ж набору значень. Із цієї ж причини не можна зберігати біометричний шаблон у зашифрованій формі (наприклад, за допомогою стан-

дартних методів шифрування, таких як RSA, AES та ін.), а потім оцінювати відповідність у зашифрованому домені.

Варто звернути увагу на те, що шифрування не є гладкою функцією і невелика різниця значень, що витягуються з початкових біометричних даних, призведе до дуже великої різниці в зашифрованому результаті. При цьому варіант з розшифровуванням шаблону і оцінюванням відповідності між збереженим та розшифрованим шаблонами, не є безпечним, оскільки має зберігатися сам біометричний шаблон. Отже, стандартні методи шифрування не є корисними для забезпечення захисту біометричних шаблонів.

Розглянемо основні схеми захисту біометричного шаблону, що на тепер набули поширення. Згідно із працями [7, 12, 14], підходи до захисту можна поділити на 2 напрями: підходи на основі перетворення властивостей та біометричні криптосистеми (рис. 3).



Рис. 3. Спрощена класифікація методів захисту шаблону

У декількох джерелах [14, 15] пропонують додаткові методи: гомоморфну криптографію, гібридні методи, а також методи на основі стеганографії та ватермаркінгу.

У підходах на основі перетворення властивостей біометричні дані обробляються за допомогою деякої функції-перетворення і далі зберігається лише вже трансформований шаблон. Залежно від типу функції-перетворення цей підхід поділяється на методи «соління» та однобічні перетворення.

У методах на основі «соління» функція-перетворення може бути оберненою [12], тобто, якщо ключ перетворення відомий по трансформованому шаблону, можна відтворити оригінальний. Безпека таких систем базується на захищеності ключа чи пароля. У методах на основі однобічних перетворень зазвичай обчислювально важко відновити оригінальний шаблон за трансформованим, навіть якщо ключ відомий [7].

Біометричні криптосистеми [1] у свою чергу поділяються на системи зі звільненням ключа (key release cryptosystems), системи зі зв'язуванням ключа (key binding cryptosystems) та системи з генерацією ключа (key generation cryptosystems).

У біометричних криптосистемах користувачеві не потрібно запам'ятовувати паролі та/або використовувати додаткові пристрої для

зберігання, передавання та ін. Біометрична криптосистема в будь-який час і в будь-якому місці ініціалізується шляхом вилучення «на льоту» необхідних параметрів з наданих біометричних зображень (з можливими помилками, стиранням тощо) без шкоди для цих зображень.

## АНАЛІЗ ЕФЕКТИВНОСТІ ІСНУЮЧИХ МЕТОДІВ ЗАХИСТУ БІОМЕТРИЧНОГО ШАБЛОНУ

Розглянемо наведені вище методи й оцінимо їх переваги та недоліки. За результатами аналізу сформуємо набір правил для вибору найкращого рішення в кожній окремій ситуації.

Основні методи захисту шаблону такі [7, 12, 13]:

*Методи «соління».* Соління або біохеш — це підхід захисту шаблону, у якому біометричні ознаки перетворюються за допомогою функції, визначеної специфічним ключем або паролем користувача. Оскільки трансформація може бути обернена, то ключ повинен бути надійно збережений користувачем та поданий під час автентифікації. Ця потреба в додатковій інформації у вигляді ключа збільшує ентропію біометричного шаблону і, отже, ускладнює для противника вгадування шаблону.

Можна відзначити переваги методу «соління». По-перше, це ефективний метод перетворення вхідних біометричних даних у високоентропійні за рахунок збільшення ентропії біометричних даних з накладанням на біометричні зразки псевдовипадкових послідовностей. По-друге, використання ключа дозволяє збільшити відстань Хемінга між даними біометричних зразків. Схема прийняття рішення за біометричної ідентифікації повинна враховувати значення кількості бітів, які збігаються у разі порівняння біометричних зразків.

*Обмеження.* 1. Якщо специфічний ключ користувача скомпрометований, шаблон більше не є безпечним, тобто, якщо противник отримує доступ до ключа та трансформованого шаблону, він може відновити оригінальний біометричний шаблон. 2. Оскільки порівняння відбувається у перетвореному вигляді, механізм «соління» повинен бути розроблений таким чином, щоб продуктивність розпізнавання не погіршувалася, навіть під час змін у біометричних даних користувача.

Використання нечітких контейнерів на основі застосування методів «соління» є ефективним методом побудови множини подань біометричних даних біометричного зразка.

*Методи на основі однобічних перетворень.* За цього підходу біометричний шаблон шифрується за допомогою однобічної функції перетворення. Параметри функції перетворення визначаються ключем, який повинен бути доступним під час автентифікації. Основною характеристикою такого підходу є те, що навіть якщо ключ та/або трансформований шаблон відомі, то обчисленням важко (через складність грубої сили) для противника відновити оригінальний біометричний шаблон.

Перевагою методу є те, що навіть якщо ключ скомпрометований, ця схема забезпечує кращу безпеку, ніж метод «соління». Заміна шаблону та ануляція можуть бути реалізовані за допомогою специфічних функцій.

*Обмеження.* Основним недоліком такого підходу є компроміс між невідповідністю та однобічністю функції перетворення. Функція перетво-

рення, з одного боку, має зберігати подібність (функції одного користувача повинні мати високу подібність у перетвореному просторі, а функції різних користувачів бути досить різноманітними після трансформації), а з другого боку, повинна бути однобічною. Важко спроектувати функції перетворення, які одночасно задовольняють обидві умови. Крім того, функція перетворення також залежить від біометричних ознак, які потрібно використовувати у певному застосуванні.

*Біометричні криптосистеми.* Традиційно, біометричні криптосистеми на нечітких екстракторах, а також системи, що їм передують, на нечітких контейнерах [2, 13], будуються з використанням завадостійкого кодування. На початковому етапі біометричні дані в певному сенсі об'єднуються з елементами завадостійких кодів (наприклад, з кодовими словами або синдромними послідовностями). Для нечітких екстракторів додатково утворюється відкритий допоміжний рядок (*helper data*), який допомагає вилучати секретний параметр на нечітких заданих біометричних даних. На етапі використання застосовується завадостійке декодування, що усуває можливу невизначеність (викликану завадами, стиранням тощо) у наданих біометричних шаблонах користувача. Якщо відмінності в наборах характеристик невеликі (не перевищують можливості коригувальних кодів), то нечіткі екстрактори (контейнери) дозволяють однозначно відновити секретний параметр (біометричний ключ).

До класу біометричних криптосистем належать три групи систем.

1. *Біометричні системи зі звільненням ключа* [1, 14]. У режимі звільнення ключа біометрична автентифікація здійснюється незалежно від механізму звільнення ключа, біометричний еталон і ключ зберігаються окремо один від одного, сам ключ звільняється після успішної біометричної автентифікації.

2. *Біометричні системи зі зв'язуванням ключа* [1, 14]. У криптографічних системах такого типу ключ і біометричний еталон криптографічно пов'язані між собою. Ключ за певним алгоритмом пов'язується з біометричним еталоном користувача і зберігається в такому вигляді в базі даних, відповідно розкрити ключ може тільки власник біометричних параметрів. У таких системах передбачається (проте не є необхідним) використання допоміжних даних (*helper data*) для демаскування зашумлених біометричних даних.

Як перевагу криптосистем цього типу слід відзначити те, що цей підхід є толерантним для змін (варіацій) даних користувача, і ця толерантність визначається здатністю коду з виправлення помилок.

Обмеження. 1. Відповідність необхідно виконати за допомогою схем корекції помилок, і це виключає використання складних схем порівняння. Це може призвести до зменшення точності порівняння. 2. Загалом біометричні криптосистеми не призначені для забезпечення різноманітності та ануляції. Проте намагаються ввести ці дві властивості в біометричні криптосистеми, головним чином, використовуючи їх у поєднанні з іншими підходами, такими як «соління». 3. Допоміжні дані мають бути ретельно зроблені.

3. *Біометричні системи з генерацією ключа* [1]. У такій біометричній криптосистемі ключ формується безпосередньо з біометричних даних кори-

стувача і не зберігається в базі даних. Варто звернути увагу на те, що якщо схема генерує той самий ключ, незалежно від шаблону вхідних даних, він має високу основну стабільність, але нульову ентропію, що зумовлює високе значення FAR. З іншого боку, якщо схема створює різні ключі для різних шаблонів того ж користувача, схема має високу ентропію, але нестабільність зумовлює високе значення FRR. Можна вивести ключ безпосередньо з біометричних ознак, однак важко одночасно досягти високої ентропії та високої стабільності.

Перевагою методу є пряма генерація ключа з біометрії.

Обмеження. Важко генерувати ключ з високою стабільністю та ентропією.

Сценарій та початкові дані відіграють важливу роль у виборі схеми захисту шаблону [7]. Наприклад, у застосуванні біометричної верифікації, такої як банкомат банку, проста схема «соління», заснована на PIN-коді користувача, може бути достатньою для забезпечення захисту біометричного шаблону. З іншого боку, під час проходження процедур аеропорту однобічне перетворення є більш придатним підходом, оскільки він забезпечує як захист шаблону, так і можливість ануляції (відкличання), не покладаючись на будь-які інші вхідні дані від користувача. Біометричні криптосистеми є більш доцільними у додатках з порівняннями на карті.

Іншим основним чинником, що впливає на вибір схеми захисту шаблону, є вибрана біометрична ознака, її набір функцій та ступінь варіацій даних користувачів. Дизайн схеми захисту шаблону залежить від конкретного типу біометрії, що використовується. Так однобічні функції були запропоновані для відбитків пальців, але важко спроектувати відповідне перетворення для райдужної оболонки ока (iris-code). Навпаки, може бути простішим розроблення біометричної криптосистеми для райдужної оболонки ока як бінарного рядка фіксованої довжини, де можна легко застосувати стандартні методи кодування з коригуванням помилок. Крім того, якщо варіації всередині даних одного типу для одного користувача досить великі, то неможливо застосувати однобічне перетворення або створити біометричну криптосистему. Тому навіть у конкретному сценарії більш ніж одна схема захисту шаблону може бути прийнятною, а вибір відповідного підходу може базуватися на ряді таких факторів, як продуктивність розпізнавання, обчислювальна складність, вимоги до пам'яті.

Оцінимо основні характеристики системи за умови використання наведених методів формування біометричного шаблону. Як характеристики системи біометричної автентифікації будемо використовувати помилки першого роду, коли визначається ймовірність помилкової відмови в доступі клієнту, який має право доступу FRR (False Rejection Rate), та помилки другого роду як ймовірність помилкового доступу, коли система помилково пізнає чужого клієнта як свого FAR (False Acceptance Rate).

## ОЦІНЮВАННЯ ЙМОВІРНОСТІ ПОМИЛКИ ПІД ЧАС ВІДДАЛЕНОЇ АВТЕНТИФІКАЦІЇ

Нехай на шаблон  $S$  довжиною  $l_S$  біт накладаються кодові слова двійкового коду  $(n, k, d)$ , що коригує помилки. При цьому під  $n$  будемо розуміти за-

гальну довжину кодів слів,  $k$  — довжина інформаційних слів і  $d$  — кодова відстань. Таких слів буде

$$N = l_s / n,$$

при цьому кількість кодів слів  $N_c = 2^k$ . Перетворення визначається операцією побітового складання слів шаблону  $S_i$  коду  $C_i$ :

$$S_i \oplus C_i = SC_i \quad i = \overline{1, N}.$$

Кодова відстань визначає можливість коригувати та визначати помилки. Код може під час декодування гарантовано виправити помилки кратністю  $t = (d - 1) / 2$  та виявити помилки кратністю  $d - 1$ .

Кодові слова генеруються за випадковими значеннями інформаційних слів  $k_i \quad i = \overline{1, N}$ . Таким чином, ключова послідовність, за якою генеруються кодові слова, повинна бути випадковою і мати довжину

$$K = kN = \frac{k}{n} l_s \text{ (бітів),}$$

де співвідношення  $R = k / n$  визначає швидкість коду.

Схема прийняття рішення порівнює зашумлені образи, які зберігаються на сервері з прийнятими з каналу зв'язку. Порівняння виконується за виразом

$$SC_i^C \oplus SC_i^K = SC_i^P, \quad i = \overline{1, N},$$

де  $N$  — кількість кодів слів, що зашумлять біометричні образи;  $SC_i^C$  — зашумлений біометричний образ, що зберігається на сервері;  $SC_i^K$  — зашумлений біометричний образ, що перевіряється на сервері.

Під час порівняння отримаємо результат

$$SC_i^P = S_i^K \oplus C_i^K \oplus S_i^C \oplus C_i^C = (S_i^K S_i^C) \oplus (C_i^K C_i^C),$$

де  $C_i^C$  — кодова послідовність, що зашумляє біометричний образ, який зберігається на сервері;  $C_i^K$  — кодова послідовність, що зашумляє біометричний образ, який перевіряється на сервері.

Сума кодів слів, які надійшли з каналу зв'язку та зашумлені в біометричних шаблонах на сервері, дає кодові слова лінійного блокового коду  $(n, k, d)$ . Отримаємо

$$SC_i^P = (S_i^K S_i^C) \oplus C_i^P, \quad i = \overline{1, N},$$

де  $C_i^P = C_i^{KC \oplus K}$  — кодові слова лінійного блокового коду  $(n, k, d)$ .

Імовірність того, що під час декодування виникнуть помилки (які будуть визначатися як розбіжність  $S_i^K$  і  $S_i^C$ ) визначається

$$p_e = 1 - \frac{N_c}{N_b}.$$

Під час автентифікації кількість помилкових кодів слів не повинна перевищувати значення порога  $T$  (дозволена для коректної роботи кількість помилок). У такому випадку отримуємо оцінку FAR імовірності помилкового доступу, коли система помилково пізнає чужого як свого:

$$FAR = \sum_{j=0}^T C_N^j p_e^j (1 - p_e)^{N-j}.$$

Декодування кодів слів коду  $(n, k, d)$  дозволяє гарантовано виявити помилки кратністю  $d - 1$ . Таким чином, код виявляє помилки з імовірністю

$$p_{i_d} = \sum_{j=1}^{d-1} C_n^j p_d^j (1 - p_d)^{n-j}.$$

Використовуючи цей вираз, отримуємо оцінку помилки першого роду FRR:

$$FRR = 1 - \sum_{j=1}^T C_N^j p_{i_d}^j (1 - p_{i_d})^{n-j}.$$

### ЗАСТОСУВАННЯ МЕТОДІВ СТЕГANOГРАФІЇ ДЛЯ ПІДВИЩЕННЯ ЗАХИЩЕНОСТІ ПРОЦЕСУ АВТЕНТИФІКАЦІЇ

Для підвищення захищеності (шляхом збільшення прихованості) біометричних даних можливе використання різних методів мережевої стеганографії, які дозволяють приховати сам факт передавання даних, необхідних для автентифікації мережею. Для застосування використано методи на основі вбудовування даних у поля мережевих протоколів.

Запропоновану схему захисту процесу віддаленої біометричної автентифікації з додаванням етапів вбудовування та завадостійкого кодування показано на рис. 4. Блоки, що відрізняють її від початкової схеми, наведені затемненим кольором.

Як показано у праці [11], для задач мережевої автентифікації найкращі показники забезпечують такі методи мережевої стеганографії: HCCUPS (0,21); TranSteg, LACK (0,18); RSTEG (0,15) та SCTP (0,12), де в дужках наведено значення векторів пріоритету.

Метод HCCUPS забезпечує найвищий рівень прихованості у зашумлених каналах, оскільки виконує маскування інформації під «природні» завади. Метод TranSteg використовується для приховування даних в IP телефонії, а також для передавання потокового відео. Для приховування інформації даний метод стискає корисне навантаження мережевого пакета за рахунок перекодування голосових даних з мінімальною втратою якості голосу і на місце, що звільнилось, в область корисного навантаження пакета вносять стеганограму; відповідно цей метод ефективний під час активної голосової чи відео сесії. Метод LACK також використовує активну RTP сесію, але його принцип дії заснований на внесенні затримки для відправлення певних голосових пакетів, корисне навантаження яких замінено. Метод RSTEG ґрунтується на повторному пересиланні пакетів і його використання для найбільшої прихованості також рекомендовано в каналах зв'язку з низьким співвідношенням сигнал/шум.



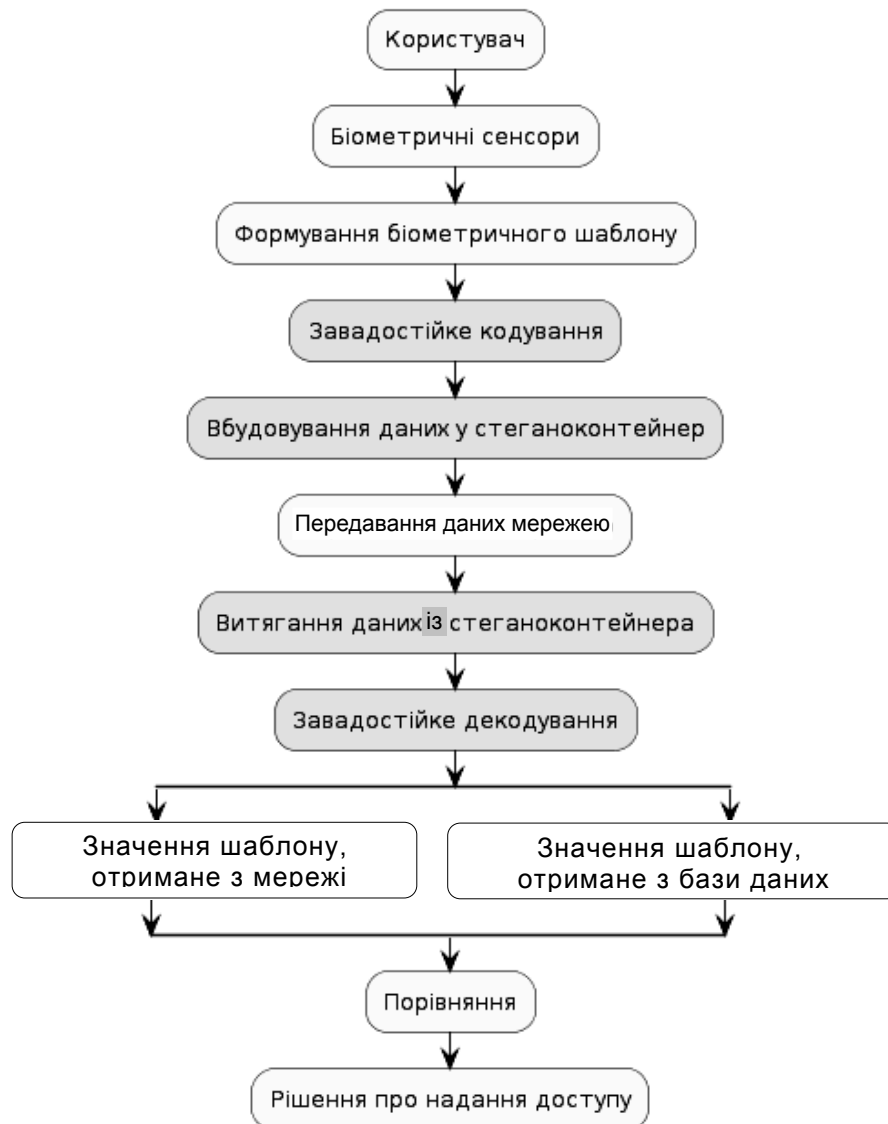


Рис. 4. Модифікована схема захисту віддаленої біометричної автентифікації

Для остаточного визначення найбільш ефективних методів вбудовування та покращення роботи системи віддаленої автентифікації в цілому побудуємо інтелектуальну систему, що буде ідентифікувати умови на вході, визначати поточний сценарій та обирати на основі правил для нього набір методів захисту.

#### **ЗАСТОСУВАННЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ПІДВИЩЕННЯ ЗАХИЩЕНОСТІ СИСТЕМИ ВІДДАЛЕНОЇ АВТЕНТИФІКАЦІЇ**

Розглянемо запропоновану систему прийняття рішень (рис. 5). Нехай на її вхід подається набір початкових даних (блок «Вхідні дані», рис. 5), який для

досліджуваного випадку містить інформацію про стан каналу зв'язку та інформацію про характер отриманих біометричних ознак користувача.



Рис. 5. Узагальнена схема прийняття рішень

*Інформація про стан каналу* [10] включає такі параметри (дослідження проводилися на прикладі LTE/5G мережі):

- параметри потужності сигналу та якості: потужність сигналу RSRP (Reference Signal Receive Power), якість сигналу RSRQ (Reference Signal Received Quality), співвідношення сигнал/шум SINR (Signal-to-Interference-plus-Noise Ratio), доступна пропускна здатність Cell Bandwidth, використовувана схема модуляції та кодування MCS (Modulation and Coding Scheme);
- наявність фонових сесій за протоколами IP, TCP, RTP, SCTP та ін.

На основі пропускної здатності, співвідношення сигнал/шум та параметрів якості/потужності визначається гранична кількість повторно переданих пакетів, що дозволяє задати пропускну здатність стеганографічних методів, які використовують повторне передавання (RSTEG, HCCUPS).

Наявність активних сесій аналізується за допомогою програмного забезпечення на пристрої користувача і може впливати на вибір методу, що використовує заміну певних полів у заголовку.

Для навчання мережі підготовлена вибірка мережевих станів, отримана з телефонів Samsung Galaxy S21, яка містить стани каналу від RSRP = -0 дБм до -120 дБм, якість RSRQ змінюється в діапазоні від -5 до -18 дБ, пропускна здатність: 10–15 МГц. На цьому етапі проводиться підготовка тестової вибірки до навчання та аналізу ефективності різних типів нейронних мереж для оброблення тестових даних.

*Інформація про характер отриманих біометричних ознак* користувача формується як результат роботи блоків отримання біометричних даних із сенсорів та формування біометричного шаблону (див. рис. 4) і включає такі елементи, як попереднє оброблення зображення, витягнення біометричних ознак та об'єднання біометричних ознак.

Процедури попереднього оброблення отриманих від сенсорів даних та витягнення біометричних ознак включають у себе стандартні процедури роботи із зображеннями, які, наприклад, у випадку оброблення райдужної оболонки ока, включають [16] нормалізацію зображення, застосування фільтра Габора, Гауса чи Лапласа та генерацію коду райдужної оболонки.

Процедура об'єднання біометричних ознак (рис. 6) передбачає пріоритизацію отриманих біометричних ознак, додавання генератора шуму для приховування полів незаповнених ознак, процедуру перемежування і завадостійкий кодер, на параметри якого впливає описана інформація про стан каналу.



Рис. 6. Процедура об'єднання біометричних ознак

Перед початком роботи системи мають бути сформовані всі дозволені сценарії роботи. Множина сценаріїв має зберігатися у спеціальній базі знань. Архітектура бази даних включає всі зазначені поля стану каналу і поле, що містить сформовані біометричні ознаки користувача. Також до неї включені поле з переліком можливих методів захисту та поле з набором дозволених алгоритмів. Обрання того чи іншого сценарію відбуватиметься на основі навчання відповідної нейронної мережі. Для недопущення випадку DoS (Denial of Service) атаки в базі знань має бути прописаний «найгірший» сценарій, який буде працювати у будь-яких умовах, але можливо з гіршими характеристиками (нижча швидкість та прихованість, більша надмірність шаблону).

Використовуючи вхідні дані, система (див. рис. 5), обирає згідно із заданими заздалегідь критеріями відповідності сценаріїв, що максимально відповідає поточному стану. Після цього з бази знань обирається набір доступних для цього сценарію методів захисту.

Приклад роботи системи і формування рішення подано на рис. 7. У цьому прикладі маємо низький рівень потужності за низької якості, що згідно із працею [17] буде відповідати параметру якості каналу (CQI) 7-9. Це дозволить обрати алгоритм модуляції 16QAM і швидкості кодування 1/3; також немає інформації про додаткові сесії, відповідно має бути обраний стійкий до завад алгоритм захисту — біометрична система зі зв'язуванням ключа та алгоритм приховування RSTEG або HICCUPS.

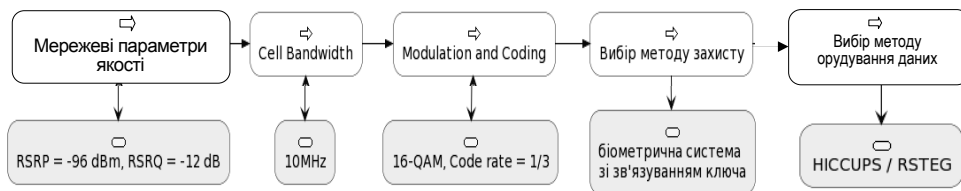


Рис. 7. Приклади роботи інтелектуальної системи прийняття рішень

## ВИСНОВКИ

У роботі вирішено актуальне завдання підвищення захищеності процесу віддаленої біометричної автентифікації шляхом покращення захищеності системи віддаленої автентифікації та застосування інтелектуальної системи прийняття рішень. Запропонована система дозволяє на основі інформації про наявні біометричні ознаки користувача обрати найкращий за заданими критеріями спосіб їх перетворення в захищений біометричний шаблон. На основі інформації про параметри каналу зв'язку запропонована система дозволяє обрати метод підвищення прихованості шляхом використання мережевої стеганографії.

У роботі запропоновано новий підхід для підвищення захищеності шляхом додавання завадостійкого кодування та використання методів мережевої стеганографії. Визначено основні атаки на систему та запропоновано методи мінімізації ймовірності їх реалізації. Проаналізовано методи формування біометричного шаблону користувача та надано рекомендації щодо їх використання.

Наукова новизна роботи полягає в застосуванні інтелектуальної системи прийняття рішень для підвищення захищеності процесу віддаленої автентифікації; уперше запропоновано використовувати мережеву стеганографію для підвищення прихованості процесу віддаленої автентифікації.

Практична значущість роботи полягає в можливості використання запропонованої системи для підвищення захищеності віддаленої біометричної автентифікації користувачів у бездротових/мобільних мережах зв'язку.

## ЛІТЕРАТУРА

1. M.S. Lutsenko, O.O. Kuznetsov, D.I. Prokopovich-Tkachenko, and V.P. Zverev, "Comparative analysis of biometric cryptosystems," (in rus), *Applied radio electronics*, vol. 17, no. 3, 4, pp. 182–191, 2018.
2. A.A. Kuznetsov, R.V. Sergienko, and A.A. Uvarova, "Fuzzy extractor on noise-tolerant codes for biometric cryptography," (in rus), *Radio engineering*, vol. 208, issue. 195, pp. 224–234.
3. Y. Dodis, R. Ostrovsky, L. Reyzin, and A.D. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *SIAM J. Comput.*, vol. 38, no. 1, pp. 97–139, 2008.
4. Y. Dodis, L. Reyzin, and A. Smith, *Fuzzy Extractors. A Brief Survey of Results from 2004 to 2006*. Available: <http://www.cs.bu.edu/~reyzin/papers/fuzzysurvey.pdf>
5. A. Juels and M. Sudan, "A fuzzy vault scheme," *Des. Codes Cryptography*, vol. 38, no. 2, pp. 237–257, 2006.
6. Anil K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circ. Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, 2004.
7. Anil K. Jain, K. Nandakumar, A. Nagar, "Biometric template security," *EURASIP J. Adv. in Signal Process.*, pp. 1–17, 2008.
8. M. Upmanyu, A.M. Namboodiri, K. Srinathan, and C.V. Jawahar, "Efficient biometric verification in encrypted domain," *ICB '09: Proc. of the Third Int. Conf. on Biometrics.*, pp. 899–908, 2009.
9. U. Uludag, S. Pankanti, S. Prabhakar, and Anil K. Jain, "Biometric cryptosystems: issues and challenges," *Proc. IEEE 2004*, 92(6), pp. 948–960.

10. A. Astrakhantsev, G. Liashenko, and A. Shcherbak, "Noise resistance of remote authentication via LTE network," *Information and Telecommunication Sciences*, vol. 2, pp. 38–43, 2020.
11. G. Liashenko, A. Astrakhantsev, and V. Chernikova, "Network steganography application for remote biometric user authentication," *IEEE 9th International Conference On Dependable Systems, Services And Technologies (DESSERT)*, pp. 326–330, 2018.
12. P. Jayapriya, R.R. Manimegalai, and R. Kumar Lakshmana, "A Survey on Different Techniques for Biometric Template Protection," *Journal of Internet Technology*, vol. 21, no. 5, 2020.
13. P. Poongodi and P. Betty, "A Study on Biometric Template Protection Techniques," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 7, no. 4, 2014.
14. Edwin T.L. Rampine and Cynthia H. Ngejane, "A Brief Overview of Hybrid Schemes for Biometric Fingerprint Template Security," in *Proceedings of the 2nd International Conference on Information Systems Security and Privacy (ICISSP 2016)*, pp. 340–346.
15. A. Sarkar and Binod K. Singh, "A Review on Different Biometric Template Protection Methods," *Recent Advances in Computer Science and Communications*, vol.14, issue 5, pp. 1551–1572, 2021.
16. V.G. Chernikova, A.A. Astrakhantsev, and G.Ye. Lyashenko, "Study of the characteristics of the iris biometric identification system", *Weapons systems and military equipment*, no.1 (53), pp.195–202, 2018.
17. 3GPP [TS 38.214]: NR; Physical layer procedures for data. Available: <https://www.tech-invite.com/3m38/tinv-3gpp-38-214.html>

Надійшла 23.06.2022

#### INFORMATION ON THE ARTICLE

**Andrii A. Astrakhantsev**, ORCID: 0000-0002-6664-3653, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: andrii.astrakhantsev@nure.ua

**Galyna Ye. Liashenko**, ORCID: 0000-0002-1741-9161, Kharkiv National University of Radio Electronics, Ukraine, e-mail: halyna.liashenko@nure.ua

#### DATA PROTECTION MANAGEMENT PROCESS DURING REMOTE BIOMETRIC AUTHENTICATION / A.A. Astrakhantsev, G.Ye. Liashenko

**Abstract.** Remote biometric authentication systems have recently become widespread due to the need to use common devices and make payments over the Internet. Because biometric methods are more user-friendly and now quickly replace passwords, the task of transmitting biometric information over an open network without compromising it is becoming urgent. This work aims to upgrade the remote authentication system to increase the secrecy and security of user biometric data. In order to achieve this goal, it is proposed to use the best security methods for forming biometric templates, network steganography to increase secrecy, and the introduction of an intelligent decision-making system. These improvements will increase the security and privacy of data during the remote authentication process.

**Keywords:** biometric template, remote authentication, attacks, network steganography.

## **RESOURCE SCHEDULING IN EDGE COMPUTING IOT NETWORKS USING HYBRID DEEP LEARNING ALGORITHM**

**G. VIJAYASEKARAN, M. DURAIPANDIAN**

**Abstract.** The proliferation of the Internet of Things (IoT) and wireless sensor networks enhances data communication. The demand for data communication rapidly increases, which calls the emerging edge computing paradigm. Edge computing plays a major role in IoT networks and provides computing resources close to the users. Moving the services from the cloud to users increases the communication, storage, and network features of the users. However, massive IoT networks require a large spectrum of resources for their computations. In order to attain this, resource scheduling algorithms are employed in edge computing. Statistical and machine learning-based resource scheduling algorithms have evolved in the past decade, but the performance can be improved if resource requirements are analyzed further. A deep learning-based resource scheduling in edge computing IoT networks is presented in this research work using deep bidirectional recurrent neural network (BRNN) and convolutional neural network algorithms. Before scheduling, the IoT users are categorized into clusters using a spectral clustering algorithm. The proposed model simulation analysis verifies the performance in terms of delay, response time, execution time, and resource utilization. Existing resource scheduling algorithms like a genetic algorithm (GA), Improved Particle Swarm Optimization (IPSO), and LSTM-based models are compared with the proposed model to validate the superior performances.

**Keywords:** edge computing, cloud computing, Internet of Things (IoT), resource scheduling, deep learning.

### **INTRODUCTION**

The Internet and smart devices have become indispensable elements in daily life. People depend on their smart devices for daily activities like payment, healthcare, virtual reality, games, etc. These different applications increase the resource requirements of smart devices. Cloud computing has been adopted to meet resource demands. The cost-effective cloud solutions offer numerous advantages in information technology and ensure that users receive essential computing, storage, and communication services based on their needs [1]. The tremendous applications based on IoT networks enhance the quality of life (Fig. 1). However, the high bandwidth requirement for IoT applications increases energy consumption, transmission bandwidth, and delay. Moreover, providing all the user-requested services using cloud computing is difficult since more than ten billion edge devices are deployed every day, and the rate is increasing [2]. To overcome these issues, edge computing paradigms have been introduced.

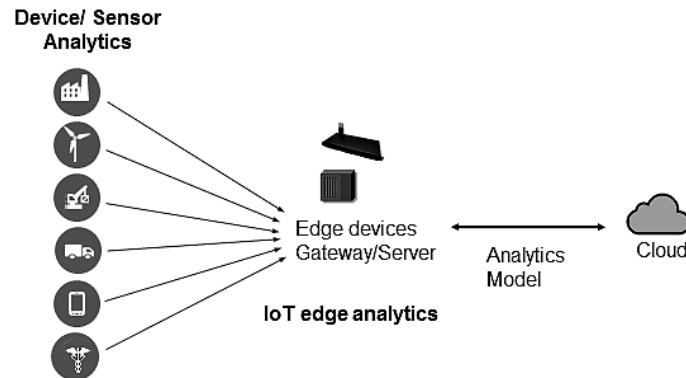


Fig. 1. IoT edge networks

Edge computing IoT networks allow the process to be performed near the device or node. The user-requested functions and services from the edge cloud are moved near to the user in edge computing to provide better storage and computing facilities [3]. Figure 1 depicts a simple illustration of IoT edge networks.

Though edge computing provides a wide range of services in various domains, it is essential to look into resource management. The increased number of user services increases the bandwidth demand on IoT networks. The resource scarcity problem and the computational complexities of IoT networks reduce the overall quality of services. Edge computing provides user-requested resources incorporating multiple techniques like clustering algorithms, and scheduling algorithms to define the user demand. Since the resource requirements to compute data types in IoT systems are different, the data processing is generally performed at regular intervals. In order to process diverse data, the system requires different computation resources, which should be provided by edge computing by switching resources from one to the other. This process will increase the computation time. To avoid this, clustering techniques are used in edge computing before allocating the resources [4].

Traditional clustering techniques like  $k$ -means, fuzzy c-means, hierarchical clustering, etc. are used in various research models. However, the conventional methods lag in performance while handling large data volumes. Moreover, conventional clustering techniques require additional dimensionality reduction techniques, which increase the overall computation cost [5]. Considering this limitation, in our previous work, we have employed an improved spectral clustering algorithm that clusters the resource requirements based on the data similarity [25]. The process flow of the clustering model is simply illustrated in Fig. 2. Compared to conventional cloud computing services, edge computing will provide minimum latency and support a wide range of IoT applications. As discussed, to satisfy the computation requirements of IoT and improve the quality of services, a hybrid deep learning-based resource allocation procedure is presented in this research work. Summarized research contributions are presented as follows.

- Presented a hybrid deep learning model for resource allocation in edge computing using deep bidirectional recurrent neural network and convolutional neural network technique.
- Presented an intense experimental analysis of the proposed model in terms of different metrics like resource utilization, response time, execution time, delay, and efficiency.

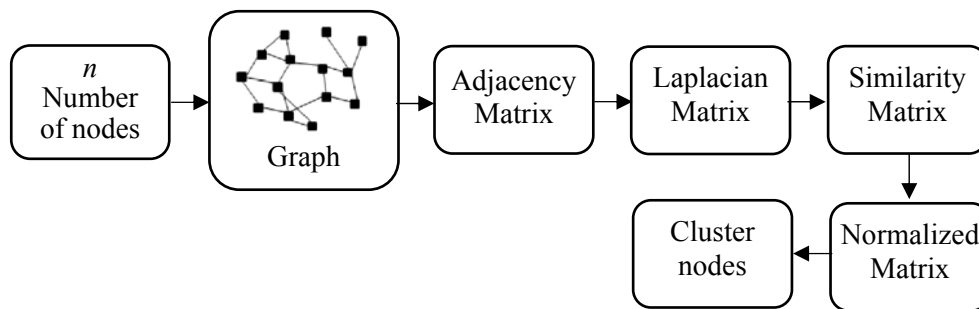


Fig. 2. Improved spectral clustering algorithm

- Presented a comparative analysis of the proposed model to validate the superior performance with conventional techniques like Improved PSO (IPSO), Genetic Algorithm (GA), and LSTM based resource scheduling procedures.

The remaining part of the article is arranged in the following order: Literature analysis of existing scheduling approaches are presented in section 2. Section 3 presents the hybrid deep learning based resource scheduling model. Section 4 presents the details of simulation results and discussion and finally, the features are concluded in section 5.

## RELATED WORKS

Resource scheduling strategies that have evolved in the past few years are considered for literature analysis, and the observations are summarized in this section based on the methodology, feature merits, and demerits. The recent trends in resource scheduling in edge computing are analyzed in [6] based on resource allocation, computation offloading, and resource provisioning. The techniques that have evolved for scheduling are categorized into centralized and distributed approaches. Applications related to these centralized and distributed approaches are analyzed and discussed in detail, which provides a basic ideology about the resource scheduling procedures. A hybrid resource scheduling procedure in edge computing was reported in [7] as a four-layer computing system that supports intelligent operations in a smart manufacturing environment. The presented two-phase hybrid algorithm incorporates greedy and threshold strategies for resource scheduling to minimize energy consumption and maximize efficiency in a manufacturing environment.

A dynamic scheduling approach for edge computing was reported in [8] incorporates deep reinforcement learning and deterministic policy gradient methods to minimize delay, energy consumption, and cache fetching costs. The presented learning models schedule the resources based on cache, offload status for uncached tasks, offloading transmission power, and edge computing resource status. The combined approach minimizes the cost function and performs better than conventional deep Q networks. The major objective of edge computing is to provide suitable computing resources for user requests in a static and dynamic environment. The issues in resource allocation are formulated as a nonlinear optimization problem in [9] and presented with a regularization-based agnostic online algorithm. The presented approach split the major issues into subcategories and formulated an objective function for each subcategory. Convex programming is



used to solve the objectives, and the experimental results validate the better performance over the online greedy one-shot solution.

A stochastic optimization was formulated in [10] considering the resource scheduling and offloading in local devices, back-end cloud and base station. To attain the objective of minimum energy consumption, and meet the QoS requirements, the stochastic optimization problem is converted into a dynamic optimization problem. To obtain an optimal solution to the dynamic optimization problem, a Lyapunov optimization theory based offloading and scheduling procedure was presented, which improves the overall performance and minimizes energy consumption. Similarly, a delay based Lyapunov function was utilized in [11], [12] to minimize the scheduling delay in multi-server edge computing systems. Without traffic statistics, the computation and communication delay can be formulated using the function and directly minimize the latency compared to traditional approaches.

Task offloading and resource scheduling in hybrid edge cloud computing reported in [13] considers the applications as graphs and analyzes the resource requirements to minimize the rent cost, time, and energy. Semidefinite and dual decomposition methods are employed to obtain offloading and scheduling decisions. Furthermore, the deep reinforcement learning model is employed to obtain discrete offloading decisions and computation frequencies. Better convergence and scheduling performance are the observed features of the presented work. The task offloading and resource scheduling procedure reported in [14] employs a decomposition model to reduce the computation complexity of the system. Logic-based bender decomposition is employed to obtain the optimal solution for master and subcategorized issues in edge computing resource management. The presented model attains better performance in delay-sensitive applications compared to conventional approaches.

An optimal task offloading and scheduling process reported in [15] considers the completion latency and energy consumption to frame the research objective. Based on the Markov decision procedure, a reinforcement learning model is employed in the presented work. Initially, it converts the problems of dynamic network conditions and task generations into a decision process. Then a double Q network is utilized along with a neural network to define the rewards attained by the system. The presented approach additionally includes a context-aware attention mechanism that assigns different weights to each action to validate better performances. Mobile edge computing, proximity-aware task offloading and scheduling were presented in [16]. The research model initially considers the distributed resources, user mobility, energy requirements, and task properties as a mixed integer non-linear programming problem. The optimal solution for the given problem is obtained using a genetic algorithm and a heuristic mobility-aware scheduling scheme was presented for effective task assignment with minimum delay and energy constraints compared to traditional methodologies. A deadline-aware task dispatching and scheduling model presented in [17] is used to schedule the new tasks and take the decision to replace the existing tasks to meet the deadline constraints. The non-trivial analysis provides better scalability compared to centralized algorithms. The presented approach minimizes the latency in sensitive applications and reduces the deadline miss ratio compared to traditional approaches.

Collaborative offloading and resource allocation algorithms reported in [18] improve the overall performance of edge computing systems and make sure that the response time limits are met. Presented migrating birds optimization algorithm identifies the optimal solution for the resource allocation problem considering the memory, CPU, energy, task queue, and servers. Maximizing the service rate with better load balancing and minimizing energy requirements are the observed features of the presented research model. A similar collaborative task scheduling model for edge computing IoT networks was reported in [19], which defines the offload state based on energy consumption and execution time. The presented approach defines when to execute the offload tasks based on the local task execution, which improves the overall throughput and deadline satisfaction ratio for critical tasks.

The resource allocation procedure for vehicle-mounted edge computing reported in [20] employed a piecewise linear approximation and relaxation procedure to obtain the optimal solution. Further, a gap-adjusted branch and bound algorithm are presented that includes a lookahead branch scheme to improve the scheduling performance over conventional scheduling models. A similar vehicular edge computing model reported in [21] incorporated the Markov decision process into the deep reinforcement learning model to obtain better training efficiency. The presented deep reinforcement learning model implementation is defined based on the proximal policy optimization algorithm. A convolutional neural network was also incorporated to approximate the value and policy functions to extract the essential features. From the literature analysis, it can be observed that the features of deep learning algorithms are not effectively utilized in resource scheduling. Most of the scheduling procedures still follow the linear and non-linear optimization problem, and solution practices. Moreover, the performance of such optimization models is also not up to the mark. Reinforcement learning is widely used in resource scheduling, but the architecture can lead to an overload state and diminish the results. Considering these limitations, a hybrid deep learning-based resource scheduling model is presented in this research work in the following section.

## **PROPOSED WORK**

The proposed resource scheduling model is developed using a bidirectional recurrent neural network model which includes a 1D-convolutional neural network model, recurrent network, and fully connected neural network block. Instead of a conventional recurrent neural network, the presented model is combined with a convolutional neural network to obtain better performances. The included convolution block is trained to learn the features from input data. The resource requirements are considered as input data and details of resources are formulated as complete information using a one-hot encoding procedure. Then the time series resource requests are processed using a bidirectional recurrent neural network model which includes a long short-term memory unit. Finally, the fully connected neural network is used to sample the output which improves the scheduling performances.

Initially, the convolutional neural network receives the requested resource details as input which includes the features like resource category, duration, sub-

class, etc., Improved spectral clustering algorithm which is utilized in our previous work is incorporated in this model also to cluster the resource requirements. These clustered information features are converted into quantified information using the encoding procedure. Instead of an integer encoding procedure, one-hot encoding is used in the proposed work. Since in integer encoding if the requirements are encoded in natural order there may be a chance for features imposed which affect the performances. We consider the time duration between resource requests as a distance factor in the encoding procedure and in the proposed work distance between each request is considered as same. i.e., the time duration between one request to another request is considered as same to reduce the experimental computation complexity.

One-hot encoding present the information as a one-dimensional array function with equal length where the value 1 in the array represents the specific resource request. Though the dimension of one hot encoding array is high due to the zero elements in the array which is used to fill the complete data. To manage this dimensionality issue, 1D-Convolutional neural networks are incorporated in the proposed architecture. CNN reduces the computational complexity by extracting the features using feature weights and activation functions. In the presented CNN model, rectified linear unit (ReLU) activation function is used and the procedure is formulated using filters  $w = \{w_1, w_2, \dots, w_\beta\}$  and features  $f = \{f_{(1)}, f_{(2)}, \dots, f_{(n)}\}$  as follows.

$$c_m = f_{(n+2p-m)} \times w_1 + \dots + f_{(n+2p)} \times w_\beta,$$

$$\text{where } m = \frac{n + 2p - \beta}{s} + 1;$$

$$f(x) = x^+ = \max(0, x),$$

where  $c_m$  is the convoluted value and the input size is reduced into  $m$  from  $n$  in one layer. if more number of layers are included then the size can be reduced further. Though employing CNN in the scheduling process reduces the computation cost significantly it reduces the complexities of recurrent block which includes LSTM and fully connected neural network models.

Following the 1D-CNN model, a bidirectional recurrent neural network is employed in the proposed architecture. The presented deep RNN model is the major element in the scheduling process that analyzes the resource requests effectively. Since the conventional deep neural network doesn't have the ability to preserve the learned information, it is not able to schedule the resource in the future. The process must be repeated again in order to satisfy the resource requests. These limitations are overcome by the presented RNN model, which learns the information in both directions, i.e., forward and backward, so that the system needs not be trained again for new requests. The RNN architecture is able to address the long-term dependencies. However, if the request gap is too large for the resources, then the RNN will exhibit poor performance. So, a type of RNN that performs better than conventional RNNs is employed in the proposed work.

Long short-term memory (LSTM) is a type of RNN model that is utilized as a bidirectional model to predict the user resource requests in the scheduling process. To attain better performance, the cells in the model obtain the input from

the previous layers and time step as a two directional propagation. The proposed network model is bidirectional, so two activation functions are employed, which are represented as  $\bar{\alpha}^t$  and  $\bar{\alpha}^t$ . The only difference between the activation functions is the input and output direction. Several gates like forget, update and output gates are used in the LSTM cells to controls the information flow over time steps. Two functions like sigmoid and tanh functions are used in the system design which is mathematically formulated as

$$\sigma(x) = \frac{1}{1 + e^{-x}},$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

The inputs for the LSTM cell are obtained from the previous layers along with time step, activation and direction steps. The cell calculates the output through the activation function along with gates. Initially the forget gate hold or drop an information. If a resource request is set then LSTM is used to track the resource schedule over time slot. If the identified resource is need to be changed due to unavailability, then its associated previous stored values will also be removed as a update process. The forget gate is the main authority which decides which information has to be stored or removed from the memory. Consider the previous activation function as  $\bar{\varepsilon}^{(t-1)}$  and its current time step is  $\varphi(t)$ , then the forget gate is represented as a function

$$C_f^t = \sigma(w_f[\alpha^{t-1}, \varphi(t)] + \vartheta_f),$$

where the weights are represented as  $w_f$ , previous activation function is represented as  $\alpha^{t-1}$   $\varphi(t)$ , the current time step input is represented as  $\varphi(t)$  and forget bias term is represented as  $\vartheta_f$ . The above function results into a vector function in the range [0,1]. Further this forget gate values are multiplied element wise with the previous cell  $C_f^{t-1}$ . If the value of the multiplication process is obtained as zero or nearer to zero then the selected resources are removed from the schedule. If the value is near to or absolutely one then the resource will be held for further process. The tanh function will create new candidate values if an information is removed by the forget gate. New candidate selection is based on the following formulation

$$C_c^t = \tanh(w_c[\alpha^{t-1}, \varphi(t)] + \vartheta_c),$$

where  $w_c$  represents the weight function and  $\vartheta_c$  represents the bias term for tanh function. The sigmoid function in the network updates the information about new candidate. The update gate has full control over the candidate values and it will add the values to the cell state. The major process of update gate is to relate the candidate to the respective previous files. The update operation is mathematically expressed as

$$C_u^t = \sigma(w_u[\alpha^{t-1}, \varphi(t)] + \vartheta_u),$$

where  $C_u^t$  is the vector and the values are present between  $[0,1]$ . Similar to previous operation an element wise multiplication is performed with  $\hat{C}^t$  to compute  $C^t$ .  $w_u$  represents the update gate weight and  $\vartheta_u$  represents the update term bias function. The information flow is decided by the forget gate and update gate decides which candidate values are need to be updated. based on this process, the new cell state is updated as follows.

$$C^t = C_f^t \times C_f^{t-1} + C_u^t \times \hat{C}^t .$$

Finally, the outputs are obtained by the output gate which considers the cell state as follows:

$$C_o^t = \sigma(w_o[\alpha^{t-1}, \varphi(t)] + \vartheta_o) ;$$

$$\alpha^t = C_o^t \times \tan h (C^t) ,$$

where the output weight is represented as  $w_o$  and bias term for output is represented as  $\vartheta_o$ . In the proposed method, the output can be the requested resource for schedule and subsequent resource requests which are need to be scheduled. Fig. 3 depicts the overall architecture of proposed resource scheduling model.

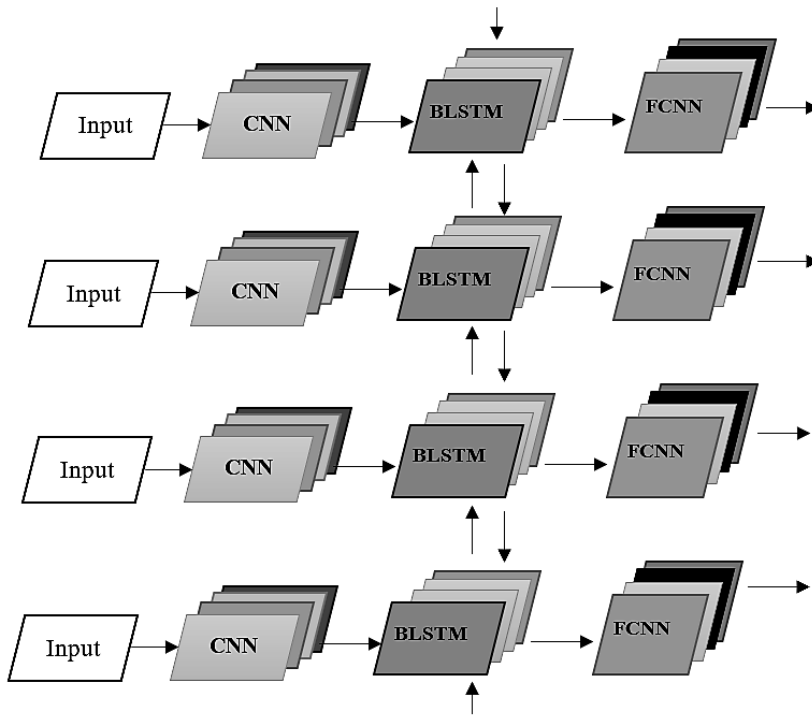


Fig. 3. Overall architecture of the proposed model

The final block in the proposed architecture is fully connected neural network block which is normal neural network. without neural network, SoftMax function can be used to predict the schedule as a probability function. The resource with low probability values are removed and high probability values are held in the scheduling process. while using fully connected network, instead of

static threshold-based result prediction, the network model learns the information from output of previous blocks. Leveraging all the outputs from bidirectional RNN the performance of proposed model is improved due to the neural network learning procedure. The process is similar to image classification from the outputs of CNN blocks but here it needs to learn from the RNN block. Mathematically the network model is formulated for one layer with activation function is given as

$$\mathcal{A}^{[\mathcal{L}]} = \sigma(\mathcal{Z}^{[\mathcal{L}]}) = \sigma(w^{[\mathcal{L}]} \mathcal{A}^{[\mathcal{L}-1]} + \mathcal{B}^{[\mathcal{L}]}) ,$$

where the activation function is represented as  $\mathcal{A}^{[\mathcal{L}]}$ , weight function is represented as  $w^{[\mathcal{L}]}$  and layer is represented as  $\mathcal{L}$ . The bias term for the layer  $\mathcal{L}$  is represented as  $\mathcal{B}^{[\mathcal{L}]}$ . The maximum probability obtained for the weight are formulated as

$$P(\mathcal{Y} = f | \varphi) = \frac{e^{\varphi w_j}}{\sum_{f=1}^{\mathcal{F}} e^{\varphi w_j}} ,$$

where the probability of prediction from SoftMax output is represented as  $P(\mathcal{Y} = f | \varphi)$ ,  $\varphi$  represents the set input and  $w_j$  represents the weights. To train the fully connected neural network, cross entropy function is employed in the proposed work. The output of the cross-entropy function measures the error between predicted results and actual results. The major objective of the training process is to reduce the prediction error. The function is mathematically formulated as

$$\mathcal{J} = -\frac{1}{\mathcal{G}} \sum_{i=1}^{\mathcal{G}} (\mathcal{O}^{(i)} \log(\alpha^{[\mathcal{L}](i)}) + (1 - \mathcal{O}^{(i)}) \log(1 - \alpha^{[\mathcal{L}](i)})) ,$$

where the desired output function is represented as  $\mathcal{O}^{(i)}$ , training samples are represented as  $\mathcal{G}$  and activation function for layer  $\mathcal{L}$  is represented as  $\alpha^{[\mathcal{L}](i)}$ . The layer parameters are updated using Adam optimizer. The optimization function is mathematically expressed as

$$w^{[\ell]} = w^{[\ell]} - \delta \frac{\rho^t \partial w^{[\ell]}}{\sqrt{\eta^t \partial w^{[\ell]} + \xi}} ;$$

where

$$\rho^t \partial w^{[\ell]} = \frac{\rho^{t-1} \partial w^{[\ell]}}{1 - (\beta_1)^t} ,$$

and

$$\rho^{t-1} \partial w^{[\ell]} = \beta_1 \rho^{t-1} \partial w^{[\ell]} + (1 - \beta_1) \frac{\partial \mathcal{J}}{\partial w^{[\ell]}} ;$$

$$\eta^t \partial w^{[\ell]} = \frac{\eta^{t-1} \partial w^{[\ell]}}{1 - (\beta_1)^t} ,$$

and

$$\eta^{t-1} \partial w^{[\ell]} = \beta_2 \eta^{t-1} \partial w^{[\ell]} + (1 - \beta_2) \frac{\partial \mathcal{J}}{\partial w^{[\ell]}} ,$$

where  $\beta_1$  and  $\beta_2$  represents the weighted average functions, the learning rate is represented as  $\delta$ , the squares of gradients before bias correction is represented as

$\eta^{t-1}$ ,  $\eta^t$  represents the after bias correction. In order to avoid zero  $\xi$  is added in the denominator. Similarly, the CNN blocks are also optimized in the samemanner as neural network block. Summarized pseudocode for the proposed resource scheduling model is presented as follows.

**Pseudocode for the proposed deep bidirectional RNN based resource scheduling**

```
Initialize learning procedure
Input: encoded data, network parameters
Output: optimal resource
While do
  Forward model
    Load network parameters and data
    For  $t \in$  time slots do
      Initialize forward propagation from CNN
        Block LSTM to neural network block
      Store the prediction results
        Store the back propagation values
      perform predicted results update
        Model backward
        For  $t \in$  time slots do
          Obtain parameter gradients in neural network block
          Obtain LSTM block parameters
          Obtain CNN block parameters
          Compute the gradient and store
          Perform update of parameters
            Update optimizer function
        End
      End
    End
  End
```

**RESULTS AND DISCUSSION**

The proposed resource scheduling algorithm using deep recurrent neural network and convolutional neural network performance is verified through simulation analysis and compared with existing resource scheduling techniques. Intel Berkeley research laboratory benchmark dataset has been used for experimentation. The dataset includes sensor readings from 54 sensors acquired from light, voltage, humidity and temperature sensors. The time duration for data collection is about 2 months and measurements are performed for every 31 seconds. Simulation analysis is performed in NetBeans version 8.1 installed in an Intel i5 processor with 16GB memory. Performance metrics like response time, execution time, resource

utilization, efficiency is considered for scheduling model. Clustering accuracy and convergence rate are used for improved spectral clustering model performance analysis. As detailed analysis of spectral clustering performances are explored in the previous research work this experimental analysis mainly focused the performance of deep learning-based scheduling techniques. Table 1 depicts the details of clustering model performance over existing *k*-means and fuzzy *c*-means clustering performances.

**Table 1.** Performance analysis of clustering models

Algorithms	Convergence Rate, %	Clustering Accuracy, %
<i>k</i> -means	94.30	92.73
FCM	95.90	95.09
Improved spectral clustering algorithm	99.00	99.15

Fig. 4 depicts the performance comparative analysis of proposed model and existing scheduling methods like improved particle swarm optimization (IPSO), genetic algorithm (GA), LSTM based scheduling methods. The resource utilization is measured based on number of clusters and maximum resource utilization is obtained by the presented model on contrary to existing methodologies. The average resource utilization attained by the IPSO and GA models are 95.92% and 95.62% which is 4% lesser than the proposed scheduling model. Resource utilization attained by the LSTM based scheduling model is 98.68% which is lesser than the proposed deep BRNN model.

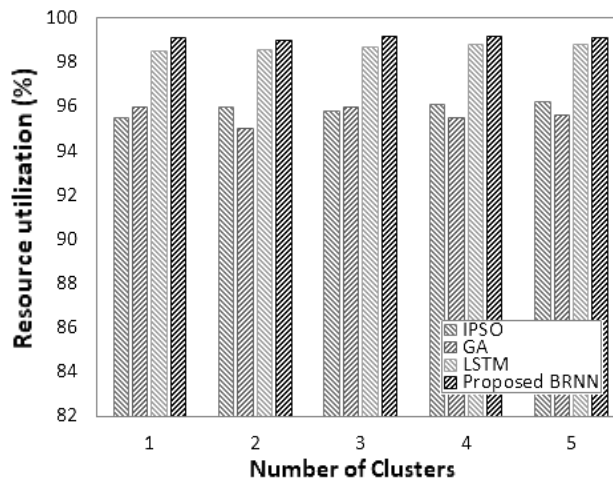


Fig. 4. Resource Utilization

The response time analysis is comparatively presented in Fig. 5 for the proposed model and existing models. The minimum time taken by the algorithm to schedule a resource is measured as response time. It is observed from the results if the number of clusters are minimum the response time of all the models is less whereas it gradually increases as the number of clusters increases. The average response acquired by the hybrid deep learning BRNN model is 1.54 seconds which is much better than the existing resource scheduling approaches. Further the performance of all the models are measured in terms of overall execution time. The process of requesting resources, time taken to check the



resource availability, time acquired to schedule the resource are collectively measured as overall execution time.

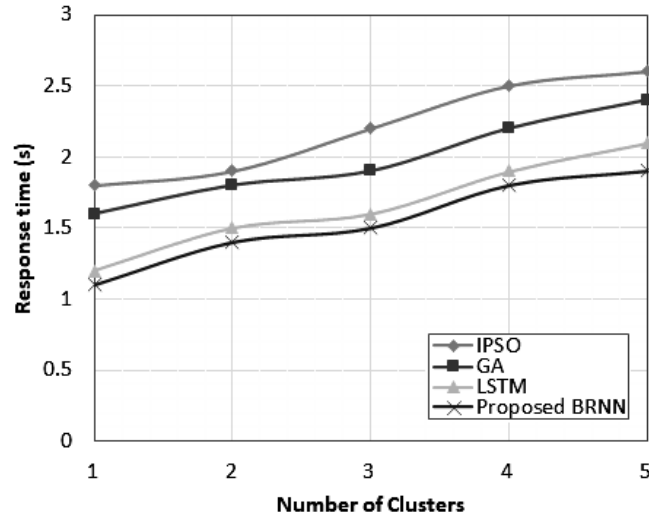


Fig. 5. Response Time Analysis

Fig. 6 depicts the comparative analysis of proposed model and existing models execution times with respect to number of iterations. The maximum iteration is selected into 50 and the execution time gradually increases for all the methods when iteration increases. The overall execution time attained by the proposed model is 14.82 seconds. Though the performance of existing methods are varied in seconds, the small time difference will introduce huge impact in service violations and affect the quality of services. Presented model schedules the resources effectively based on the prediction characteristics of bidirectional model which reduces the further analysis if same resource is requested in future. Whereas existing methods analyze the request again and confirms the resource status and schedule to the respective request increases the overall execution time.

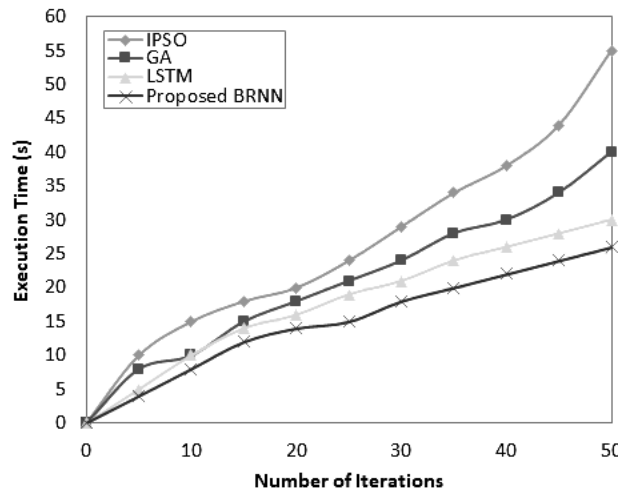


Fig. 6. Execution Time Analysis

The average delay exhibited by the existing models and proposed model is analyzed and depicted in Fig. 7. The time taken by a user to acquire an optimal resource is generally termed as waiting time. Whereas delay describes about the

time which exceeds the stipulated predefined time period. Since all the requests are fixed with a minimum waiting time and meanwhile the system need to search for the resource and schedule them for further process. if the minimum waiting time is exceeding over a period and the resources are scheduled after that then it is measured as delay. Here the proposed model exhibits minimum delay compared to existing scheduling procedures. The prediction performance and memory utilization of proposed model not only reduces the execution time also it reduces the delay by schedule the resource based on existing utilization. Whereas there is no process followed in the existing scheduling procedures increases the delay.

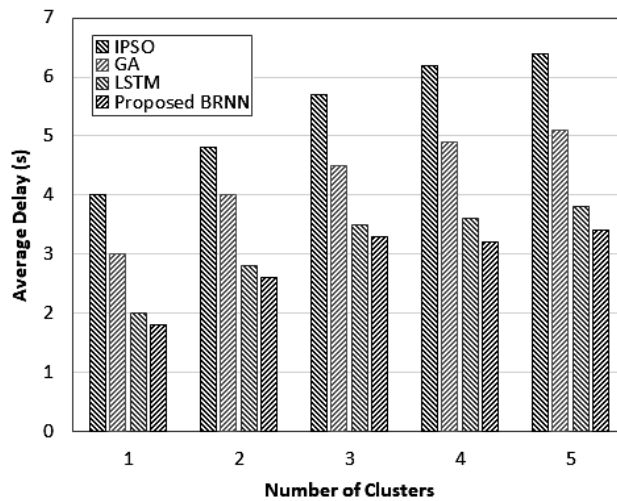


Fig. 7. Average Delay Analysis

The overall efficiency of all the approaches are comparatively analyzed and depicted in Fig. 8. Based on the response time, resource utilization, execution time and delay the efficiency is measured. If the algorithm exhibits maximum delay definitely it will introduce an impact in the efficiency. Similarly, if the resource utilization is low then that system could not be considered as an efficient one. It is essential for a system to complete the required process with minimum response time and execution time. Considering all these factors, the proposed model attains maximum efficiency score compared to other methods. Since the proposed scheduling procedure reduces the response time, execution time and improves the resource utilization and minimizes the delay which indicates the maximum efficiency.

Table 2. Performance Comparative Analysis

Methods	Resource Utilization (%)	Response Time (s)	Execution Time (s)	Average Delay (s)	Efficiency (%)
IPSO	95.92	2.20	26.09	5.42	94.50
GA	95.62	1.98	20.73	4.30	96.00
LSTM	98.68	1.66	17.55	3.14	98.00
Proposed BRNN	99.12	1.54	14.82	2.86	98.90

Table 2 depicts the summary of proposed model and existing resource scheduling models performances in terms of resource utilization, response time,

execution time, average delay and efficiency. It can be observed from the results that the proposed model attains better performance than existing approaches. improved scheduling performance will increase the overall performance and quality of services in cloud integrated IoT networks.

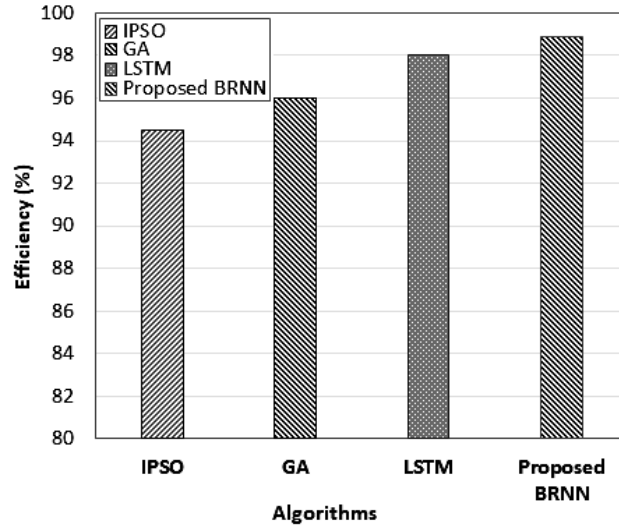


Fig. 8. Efficiency Analysis

## CONCLUSION

A hybrid deep learning model for resource scheduling in edge computing Internet of Things (IoT) network is presented in this research work. The presented scheduling algorithm includes the deep bidirectional recurrent neural network with convolutional neural network block to improve the scheduling performance in edge computing. Initially the resource requests are clustered using improved spectral clustering algorithm and converted into encoded image. The encoded information features are processed by one-dimensional convolutional neural network model followed by bidirectional long-short term memory which is type of RNN model. The final results select the optimal resources and schedule to respective requests to reduce the computational complexity of IoT network. Simulation analysis of proposed model demonstrates the better performances compared to existing scheduling models which is based on improved particle swarm optimization algorithm, genetic algorithm and LSTM based model. Further, this research work can be extended by introducing concatenated deep learning techniques to avoid initial clustering process and improve the overall performances.

## REFERENCES

1. Vishnu Kumar Kaliappan, S. Gnanamurthy, and K. Mohanasundaram, "Reduced power consumption by resource scheduling in mobile cloud using optimized neural network," *Materials Today: Proceedings*, vol. 46, no.15, pp.6453–6458, 2021.
2. Yanling Shao, Chunlin Li, and Youlong Luo, "Cost-effective replication management and scheduling in edge computing," *Journal of Network and Computer Applications*, vol.129, pp. 46–61, 2019.

3. Xiao Ma, Shangguang Wang, Shan Zhang, Peng Yang, Chuang Lin, and Xuemin Shen, "Cost-Efficient Resource Provisioning for Dynamic Requests in Cloud Assisted Mobile Edge Computing," *IEEE Transactions on Cloud Computing*, vol. 9, no. 3, pp. 968–980, 2021.
4. Yongsheng Hao, Jie Cao, and Jinglin Du, "Energy-aware scheduling in edge computing with a clustering method," *Future Generation Computer Systems*, vol.117, pp.259–272, 2021.
5. Jun Liu, Tianfu Yang, and Bo Sun, "Resource allocation and scheduling in the intelligent edge computing context," *Future Generation Computer Systems*, vol.121, pp. 48–53, 2021.
6. Quyuan Luo, Shihong Hu, Changle Li, Guanghui Li, and Weisong Shi, "Resource Scheduling in Edge Computing: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2131–2165, 2021.
7. Xiaomin Li, Jiafu Wan, Hong-Ning Dai, Muhammad Imran, Min Xia, and Antonio Celesti, "A Hybrid Computing Solution and Resource Scheduling Strategy for Edge Computing in Smart Manufacturing," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4225–4234, 2019.
8. Samrat Nath and Jingxian Wu, "Deep reinforcement learning for dynamic computation offloading and resource allocation in cache-assisted mobile edge computing systems," *Intelligent and Converged Networks*, vol. 1, no. 2, pp. 181–198, 2020.
9. Lin Wang, Lei Jiao, Jun Li, Julien Gedeon, and Max Mühlhäuser, "MOERA: Mobility-Agnostic Online Resource Allocation for Edge Computing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 8, pp. 1843–1856, 2019.
10. Fengjun Zhao, Ying Chen, Yongchao Zhang, Zhiyong Liu, and Xin Chen, "Dynamic Offloading and Resource Scheduling for Mobile-Edge Computing With Energy Harvesting Devices," *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 2154–2165, 2021.
11. Yuan Zhang, Peng Du, Jiang Wang, Teer Ba, Rui Ding, and Ning Xin, "Resource Scheduling for Delay Minimization in Multi-Server Cellular Edge Computing Systems," *IEEE Access*, vol. 7, pp. 86265–86273, 2019.
12. Bin Zhang and Dexu Chen, "Resource scheduling of green communication network for large sports events based on edge computing," *Computer Communications*, vol.159, pp. 299–309, 2020.
13. Qi Zhang, Lin Gui, Shichao Zhu, and Xiupu Lang, "Task Offloading and Resource Scheduling in Hybrid Edge-Cloud Networks," *IEEE Access*, vol. 9, pp. 85350–85366, 2021.
14. Hyame Assem Alameddine, Sanaa Sharafeddine, Samir Sebbah, Sara Ayoubi, and Chadi Assi, "Dynamic Task Offloading and Scheduling for Low-Latency IoT Services in Multi-Access Edge Computing," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 3, pp. 668–682, 2019.
15. Tong Liu, Yameng Zhang, Yanmin Zhu, Weiqin Tong, and Yuanyuan Yang, "Online Computation Offloading and Resource Scheduling in Mobile-Edge Computing," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6649–6664, 2021.
16. Umber Saleem, Yu Liu, Sobia Jangsher, Yong Li, and Tao Jiang "Mobility-Aware Joint Task Scheduling and Resource Allocation for Cooperative Mobile Edge Computing," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 360–374, 2021.
17. Jiaying Meng, Haisheng Tan, Xiang-Yang Li, Zhenhua Han, and Bojie Li, "Online Deadline-Aware Task Dispatching and Scheduling in Edge Computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 6, pp. 1270–1286, 2020.
18. Haitao Yuan and MengChu Zhou, "Profit-Maximized Collaborative Computation Offloading and Resource Allocation in Distributed Cloud and Edge Computing Systems," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 3, pp. 1277–1287, 2021.

19. Youngjin Kim, Chiwon Song, Hyuck Han, Hyungsoo Jung, and Sooyong Kang, "Collaborative Task Scheduling for IoT-Assisted Edge Computing," *IEEE Access*, vol. 8, pp. 216593–216606, 2020.
20. Yu Liu, Yong Li, Yong Niu, and Depeng Jin, "Joint Optimization of Path Planning and Resource Allocation in Mobile Edge Computing," *IEEE Transactions on Mobile Computing*, vol. 19, no. 9, pp. 2129–2144, 2020.
21. Wenhan Zhan et al., "Deep-Reinforcement-Learning-Based Offloading Scheduling for Vehicular Edge Computing," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5449–5465, 2020.
22. Yi-Han Chiang, Tianyu Zhang, and Yusheng Ji, "Joint Cotask-Aware Offloading and Scheduling in Mobile Edge Computing Systems," *IEEE Access*, vol. 7, pp. 105008–105018, 2019.
23. Shuaishuai Guo, Dalei Wu, Haixia Zhang, and Dongfeng Yuan, "Resource Modeling and Scheduling for Mobile Edge Computing: A Service Provider's Perspective," *IEEE Access*, vol. 6, pp. 35611–35623, 2018.
24. Tien Van Do, N.H. Do, and Peter Hegyi, "Comparison of scheduling algorithms for multiple mobile computing edge clouds," *Simulation Modelling Practice and Theory*, vol. 93, pp. 104–118, 2019.
25. G. Vijayasekaran and M. Duraipandian, *An Efficient Clustering and Deep Learning Based Resource Scheduling for Edge Computing to Integrate Cloud-IoT. Wireless Pers Commun.* 2022. Available: <https://doi.org/10.1007/s11277-021-09442-8>.

Resieved 00.00.2022

#### INFORMATION ON THE ARTICLE

**G. Vijayasekaran**, Department of Computer Science and Engineering, Sir Issac Newton College of Engineering and Technology, Nagapattinam, India, e-mail: gvijayasekaran2@gmail.com

**M. Duraipandian**, Department of Computer Science and Engineering, Hindusthan Institute of Technology, Coimbatore, India, e-mail: durainithi@gmail.com

**ПЛАНУВАННЯ РЕСУРСІВ У МЕРЕЖАХ ІОТ EDGE COMPUTING З ВИКОРИСТАННЯМ ГІБРИДНОГО АЛГОРИТМУ ГЛИБОКОГО НАВЧАННЯ /**  
Г. Віджаясекаран, М. Дурайпандіан

**Анотація.** Поширення Інтернету речей (IoT) і бездротових сенсорних мереж покращує передачу даних. Попит на передачу даних швидко зростає, що викликає появу парадигми периферійних обчислень. Граничні обчислення відіграють важливу роль у мережах IoT і надають обчислювальні ресурси поблизу користувачів. Перенесення служб із хмари до користувачів розширює комунікаційні, сховища та мережеві функції користувачів. Однак масивні мережі IoT потребують великого обсягу ресурсів для своїх обчислень. Щоб досягти цього, у граничних обчисленнях використовуються алгоритми планування ресурсів. Алгоритми планування ресурсів, засновані на статистиці та машинному навчанні, розвинулися протягом останнього десятиліття, але їх продуктивність можна покращити, якщо додатково проаналізувати вимоги до ресурсів. У роботі подано глибоке планування ресурсів на основі навчання в периферійних обчислювальних мережах IoT з використанням глибокої двонаправленої рекурентної нейронної мережі (BRNN) і алгоритмів згорткової нейронної мережі. Перед плануванням користувачі IoT класифікуються в різні кластери за допомогою спектрального алгоритму кластеризації. Пропонований аналіз моделювання перевіряє продуктивність з точки зору затримки, часу відгуку, часу виконання та використання ресурсів. Існуючі алгоритми планування ресурсів, як от генетичний алгоритм (GA), покращена оптимізація роїв частинок (IPSO) і моделі на основі LSTM, порівнюються із запропонованою моделлю для підтвердження кращої продуктивності.

**Ключові слова:** периферійні обчислення, хмарні обчислення, інтернет речей (IoT), планування ресурсів, глибоке навчання.

## МЕТОДИ ТА МОДЕЛІ НЕЙРОМЕРЕЖЕВОЇ АПРОКСИМАЦІЇ ГРАДУЮВАЛЬНИХ ХАРАКТЕРИСТИК NTC-ТЕРМІСТОРІВ

С.С. ФЕДІН, І.С. ЗУБРЕЦЬКА

**Анотація.** Підтверджено гіпотезу про доцільність застосування RBF-мереж для підвищення точності побудови градуювальних характеристик NTC-термісторів у робочому діапазоні температур без поділу його на піддіапазони. Встановлено, що похибка нейромережевої апроксимації градуювальних характеристик NTC-термісторів на основі RBF-мереж не менше ніж у півтора рази нижча за допустиму похибку апроксимації поліноміальної моделі третього порядку, яка використовується в програмному забезпеченні сучасних систем збирання та оброблення вимірювальної інформації. Розроблено методику оброблення вимірювальної інформації з використанням RBF-мереж для автоматизації процедури побудови індивідуальних градуювальних характеристик і періодичного калібрування NTC-термісторів.

**Ключові слова:** точність, вимірювальна інформація, градуювальна характеристика, NTC-термістор, робочий діапазон температур, функція перетворення, нейромережева апроксимація, RBF-мережа.

### ВСТУП

Підвищення точності оброблення вимірювальної інформації та нормування метрологічних характеристик засобів вимірювальної техніки (ЗВТ) зокрема для градуювання та калібрування первинних вимірювальних перетворювачів (сенсорів), є однією із визначальних умов забезпечення єдності вимірювань.

При цьому актуальним завданням є забезпечення точності побудови градуювальних характеристик (ГХ) температурних сенсорів, оскільки, як показує аналіз вітчизняних і зарубіжних досліджень, приблизно половина вимірювань у промисловості виконується для визначення значень температури (рис. 1) [1, 2].

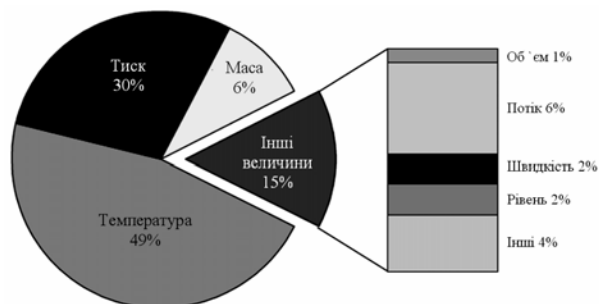


Рис. 1. Розподіл видів вимірювань фізичних величин у промисловості

У випробувальних і калібрувальних лабораторіях процедура побудови ГХ контактних сенсорів температури здійснюється з використанням сучас-

них інформаційно-вимірювальних систем збирання даних (СЗД) і потребує врахування систематичної складової похибки нелінійних функцій перетворення сенсорів, величина якої залежить від досконалості обчислювального алгоритму оброблення експериментальних даних. Зокрема під час градування напівпровідникових сенсорів із нелінійною функцією перетворення і від'ємним температурним коефіцієнтом електричного опору (NTC-термісторів) складова похибки апроксимації ГХ має бути несуттєвою порівняно з похибкою результатів вимірювання температури [3, 4].

Сучасні NTC-термістори характеризуються високою чутливістю, малою інерційністю, стійкістю до хімічних і механічних впливів, компактністю та відносно низькою вартістю. Ці властивості визначають можливість ефективного використання NTC-термісторів як чутливих елементів ЗВТ для вимірювання середніх температур у легкій, нафтовій, газовій, харчовій, фармацевтичній промисловості, у медичній та автомобільній техніці, побутовій і промисловій електроніці, сільському та комунальному господарстві. Однак застосування NTC-термісторів обмежене через значну нелінійність залежності температури від електричного опору та великий розкид їх характеристик [5]. Це потребує побудови індивідуальних ГХ у робочому діапазоні температур.

Вирішення цього завдання запропоновано Дж. Стейнхартом і Ст. Хартом на основі застосування поліноміальної моделі третього порядку

$$T^{-1} = A + B(\ln R) + C(\ln R)^3, \quad (1)$$

де  $T$  — температура, К;  $R$  — електричний опір, Ом;  $A$ ,  $B$ ,  $C$  — параметри, які визначаються під час градування у трьох вузлових точках з інтервалом не менше ніж  $10^\circ\text{C}$  за умови максимального наближення до експериментальної залежності [6].

Функція перетворення NTC-термісторів має експоненційний характер, тому для забезпечення високої точності при побудові ГХ робочий діапазон температур поділяють на окремі піддіапазони. При цьому модель (1), яка використовується в обчислювальних алгоритмах СЗД, має різні значення параметрів у кожному піддіапазоні, що зумовлює похибку функції перетворення NTC-термістора  $\Delta = 0,05^\circ\text{C}$  [7].

Проведений аналіз літературних джерел дозволив встановити, що підвищення точності побудови нелінійних ГХ NTC-термісторів у робочому діапазоні температур можливе за рахунок застосування в обчислювальних алгоритмах СЗД моделей оброблення вимірювальної інформації, створених на основі методів інтелектуального аналізу даних, до яких належить нейромережеве моделювання.

Таким чином, підвищення точності оброблення вимірювальної інформації за побудови ГХ NTC-термісторів у робочому діапазоні температур з використанням методів нейромережевого моделювання є актуальним завданням, яке має наукове та практичне значення.

## ПОСТАНОВКА ЗАДАЧІ

Нехай нелінійне відображення «вхід»  $\rightarrow$  «вихід» ГХ NTC-термісторів подано у вигляді векторної функції  $f(\cdot)$ , яка вважається невідомою й інфор-

мація про неї надається у вигляді множини  $P$  прикладів (векторів), що отримані за вхідними даними температури  $T_i$  і вихідними даними електричного опору  $R_i$

$$P = \{(T_i, R_i)\}_{i=1}^N. \quad (2)$$

Тоді нейромережеву апроксимацію ГХ NTC-термісторів можна здійснити на підставі інформації, поданої у вигляді навчальної вибірки (2). При цьому структура нейронної мережі (НМ), що апроксимує функцію  $f(\cdot)$ , повинна відповідати умові: функція  $F(\cdot)$ , яка описує відображення вхідного сигналу у вихідний, має бути досить близькою до функції  $f(\cdot)$  у вимірі евклідової норми на множині всіх вхідних векторів  $x$ , тобто

$$\|F(x) - f(x)\| < \varepsilon, \quad (3)$$

де  $\varepsilon$  — деяке мале додатне число.

Оскільки ГХ NTC-термісторів є експоненційною, то функцію відображення  $F(\cdot)$  можна отримати на основі моделі прямошарові НМ з радіально-базисними функціями (RBF), застосування яких дозволяє експоненційно зменшувати локалізовані нелінійності при апроксимації складних залежностей [8]. RBF-мережа має один прихований шар із  $m_1$  нейронами і дозволяє відобразити функцію із замкненої області  $\mathfrak{R}^{m_0}$  у простір  $\mathfrak{R}$ , де  $m_0$  — розмірність входів навчальної вибірки. Метод використання RBF-мереж, заснований на виборі функції, в якості якої в більшості випадків використовується радіальна базисна функція Гауса (GRB-функція)

$$F(x) = \sum_{i=1}^{m_1} w_i \exp\left(-\frac{\|x - t_i\|^2}{2\sigma^2}\right), \quad (4)$$

де  $w_i \in \mathfrak{R}$ ,  $t_i \in \mathfrak{R}^{m_0}$ ,  $\sigma > 0$  для  $i = 1, 2, \dots, m_1$ ,  $t_i$  — вектор координат центрів RBF [8].

У порівнянні з класичною апроксимацією і виконанням умови (3) застосування нейромережевих моделей дає змогу адаптуватися до вхідних даних з наближенням прямої  $R = f(T)$  і оберненої  $T = f^{-1}(R)$  функцій перетворення при калібруванні NTC-термісторів, які характеризуються великим розкидом значень. При цьому необхідно виконати сумісні вимірювання  $T$  і  $R$ , отримати множину (2) і подати її у вигляді навчальної вибірки для обчислювальних експериментів.

## МЕТА, ОБ'ЄКТ І ПРЕДМЕТ ДОСЛІДЖЕННЯ

Мета роботи — підвищення точності оброблення вимірювальної інформації під час побудови індивідуальних градуювальних характеристик NTC-термісторів у робочому діапазоні температур з використанням методів нейромережевого моделювання.

Об'єкт дослідження — процес забезпечення точності вимірювань температури з використанням NTC-термісторів.

Предметом дослідження є методи і моделі оброблення вимірювальної інформації для побудови градуювальних характеристик NTC-термісторів.



## ОГЛЯД ЛІТЕРАТУРИ

Забезпеченню точності ГХ ЗВТ присвячено праці О.А. Герашенка, В.А. Грановського, Р. Джексона, В.Т. Кондратова, Дж. Фрайдена [9–13]. Результати цих досліджень базуються на використанні схемотехнічних, структурно-алгоритмічних і параметричних методів градування і калібрування сенсорів.

У працях О.В. Дегтярьова, О.В. Запорожця, І.В. Руженцева, N.J. Cotton, T.D. Dongale, S.A. Khan та інших дослідників для вирішення завдання зменшення похибки апроксимації ГХ сенсорів запропоновано використання методів інтелектуального оброблення вимірювальної інформації, до яких належить нейромережеве моделювання [14–19]. Результати цих досліджень свідчать про доцільність побудови ГХ NTC-термісторів на основі моделей прямошарових НМ типу перцептрон (MLP). Такі моделі є інваріантними до будь-якого вигляду нелінійності функцій перетворення сенсорів, однак застосовуються для апроксимації ГХ NTC-термісторів, у вузькому діапазоні температур, до якого не належить діапазон вимірювання температури з використанням NTC-термісторів (- 55...155 °С).

Попередні власні дослідження та аналіз властивостей різних типів НМ показують, що для рівномірної апроксимації будь-якої неперервної функції на компактній множині значень можуть бути використані RBF-мережі [20, 21, 22]. Відповідно до ДСТУ ISO/IEC 2382:2017 RBF-мережа — прямошарова НМ, у якій кожен прихований нейрон використовує радіальну базисну функцію активації, а вихідні нейрони реалізують лінійні комбінації функцій активації нейронів прихованого шару [23].

RBF-функцією називається функція, яка радіально змінюється навколо деякого центра і набуває ненульових значень тільки в околі цього центра [24]. У праці [8] показано, що використання радіальних базисних функцій Гауса дозволяє експоненційно зменшувати локалізовані нелінійності за апроксимації складних залежностей. Як і мережі типу MLP, RBF-мережі належать до багатошарових мереж прямого поширення і є універсальними апроксиматорами. Незважаючи на те, що завжди існує RBF-мережа, яка здатна імітувати мережу MLP [8], ці типи нейромережевих моделей мають важливі характеристики, наведені в табл. 1.

**Таблиця 1.** Порівняння властивостей RBF-мереж і багатошарових перцептронів

Характеристика	Тип НМ	
	MLP	RBF-мережа
Кількість прихованих шарів	Більше одного	Один
Вигляд моделі нейрона у прихованих шарах і вихідному шарі	Однакова (одні і ті ж цілі)	Різна (різні цілі)
Лінійність і (або) нелінійність шарів	Нелінійні приховані шари і вихідний шар	Нелінійний прихований шар; вихідний шар – лінійна комбінація виходів нейронів прихованого шару
Аргумент функції активації	Скалярний добуток вхідного вектора і вектора синаптичних ваг даного нейрона	Евклідова норма (відстань) між вхідним вектором і центром радіальної базисної функції
Вид апроксимації	Глобальна апроксимація нелінійного відображення	Локальна апроксимація нелінійного відображення за рахунок локалізованих нелінійностей, що зменшуються експоненційно

Таким чином, аналіз властивостей RBF-мереж і MLP дозволяє зробити припущення про те, що підвищення точності апроксимації ГХ NTC-термісторів у робочому діапазоні температур (- 55...155 °С) без поділу його на піддіапазони можливе за рахунок використання радіально-базисних функцій, оскільки залежність (1) є експоненційно нелінійною, а радіально-базисні НМ забезпечують рівномірну апроксимацію будь-якої неперервної функції на компактній множині значень.

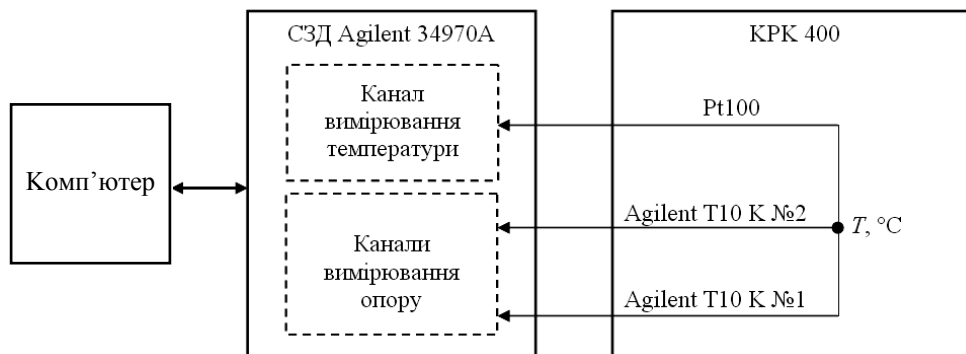
## МАТЕРІАЛИ ТА МЕТОДИ

Фізичний експеримент здійснювали за запропонованою. Результати цих досліджень методикою, у відповідності до якої в центрі випробувального відсіку кліматичної камери КРК 400 розміщували NTC-термістори Agilent T10 K і TDK B57861S (табл. 2) і опорний термометр опору Pt100 з похибкою вимірювання  $\pm 0,05$  °С [25].

**Таблиця 2.** Характеристики NTC-термісторів

Сенсори	Робочий діапазон температур, °С	Номинальний опір за температури +25 °С, кОм	Похибка, °С
Agilent T10 K	-70...70	10	$\pm 0,2$
TDK B57861S	-55...155	10	$\pm 0,2$

На першому етапі експерименту після підключення сенсора Pt100 до каналу вимірювання температури та двох NTC-термісторів Agilent T10 K до каналів вимірювання електричного опору налаштовували СЗД Agilent 34970A з використанням спеціалізованого програмного забезпечення BenchLink Data Logger 3 (рис. 2).



*Рис. 2.* Структурна схема фізичного експерименту із застосуванням двох NTC-термісторів Agilent T10 K

Систему збирання даних Agilent 34970A підключали до комп'ютера через послідовний порт RS-232 і вимірювали температуру в автоматичному режимі (рис. 3).

З урахуванням характеристик сенсорів Agilent T10 K (див. табл. 2) мінімальне значення температури встановлювали на рівні -70 °С. Температурний режим випробувань у КРК 400 програмували за такими параметрами, як діапазон температури -70...70 °С; крок зміни температури 5 °С; час зміни температури 15 хвилин; час витримання за встановленою темпера-

турою у кожній вузловій точці робочого діапазону був визначений експериментально та становив 30 хвилин. Випробування термісторів Agilent T10 K проводили з метою оцінювання розкиду функцій перетворення сенсорів на підставі розрахунку зведеної похибки з використанням відібраних даних температури в опорних точках:  $-70\text{ }^{\circ}\text{C}$ ,  $-35\text{ }^{\circ}\text{C}$ ,  $0\text{ }^{\circ}\text{C}$ ,  $35\text{ }^{\circ}\text{C}$  і  $70\text{ }^{\circ}\text{C}$ . Зведену похибку NTC-термісторів  $\Delta_{\text{NTC}}^{3\text{B}}$  розраховували за формулою

$$\Delta_{\text{NTC}}^{3\text{B}} = \left( \frac{T_{\text{NTC}} - T_{\text{Pt100}}}{T_p} \right) \cdot 100, \quad (5)$$

де  $T_{\text{NTC}}$  і  $T_{\text{Pt100}}$  — відповідно значення температури NTC-термісторів Agilent T10 K, які отримано на основі залежності (1), та опорні значення температури;  $T_p$  — протяжність робочого діапазону температури термісторів.

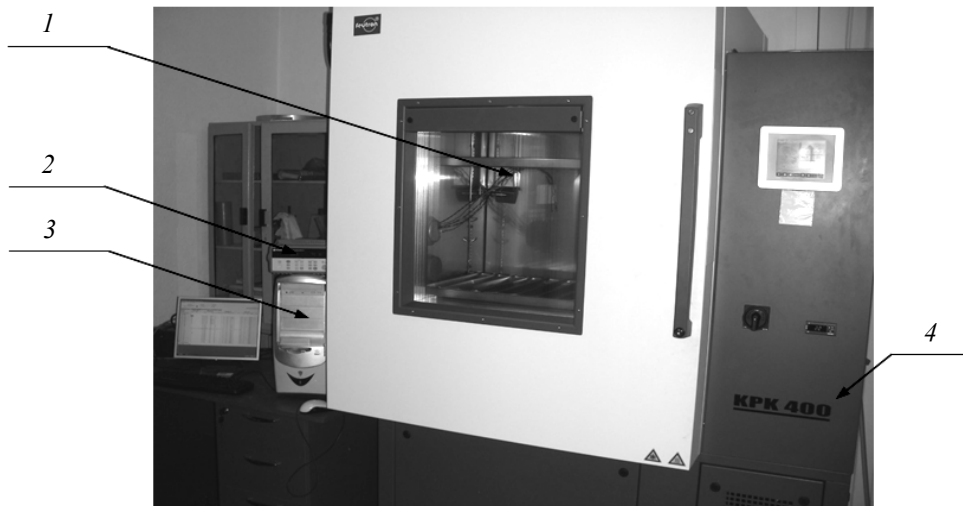


Рис. 3. Апаратна реалізація фізичного експерименту: 1 — сенсори; 2 — СЗД Agilent 34970A; 3 — комп'ютер; 4 — кліматична камера KPK 400

Результат оцінювання значущості відмінностей зведених похибок  $\Delta_{\text{NTC}\#1}^{3\text{B}}$  і  $\Delta_{\text{NTC}\#2}^{3\text{B}}$  з використанням  $t$ -тесту для незалежних змінних, отриманий в системі STATISTICA 6.1, показує відмінність зведеної похибки температури двох сенсорів Agilent T10 K на рівні  $p = 0,046$ , що підтверджує доцільність їх індивідуального градуювання (рис. 4) [25].

Т-критерий независимых выборок (Погрешности_Agilent.sta)					
Замечание: Переменные рассм. как независимые выборки					
Группа 1 и Группа 2	Среднее Группа 1	Среднее Группа 2	t-знач.	ст. св.	p
$\Delta_{\text{NTC}\#1}^{3\text{B}}$ vs. $\Delta_{\text{NTC}\#2}^{3\text{B}}$	0,114714	0,356143	-2,36039	8	0,045930

Рис. 4. Скріншот результату оцінювання значущості відмінностей зведених похибок сенсорів Agilent T10 K

На другому етапі експерименту проводили випробування шести термісторів TDK B57861S, кількість яких обумовлена мінімальним обсягом вибірки

для забезпечення достовірності статистичного оброблення отриманих результатів. Сенсор №1 підключали до каналу вимірювання електричного опору СЗД Agilent 34970A, а сенсори №2–№6 і опорний термометр опору Pt100 — до каналів вимірювання температури (рис. 5).

Під час проведення випробувань сенсорів TDK B57861S мінімальне значення температури встановлювали на рівні  $-55\text{ }^{\circ}\text{C}$  і програмували режим проведення випробувань за параметрами: діапазон температури  $-55\dots150\text{ }^{\circ}\text{C}$ , що обумовлено технічними характеристиками СЗД Agilent 34970A; крок зміни температури  $5\text{ }^{\circ}\text{C}$ ; час зміни температури — 15 хв; час витримування за встановленої температури — 30 хв.

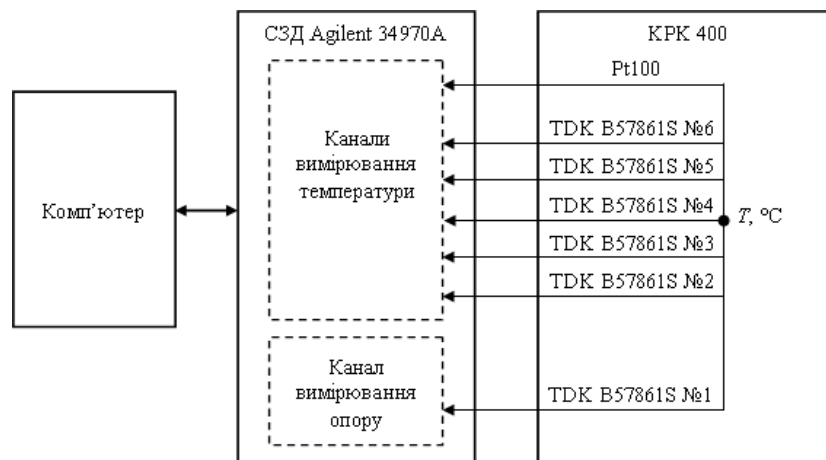


Рис. 5. Структурна схема фізичного експерименту із застосуванням шести NTC-термісторів TDK B57861S

У результаті проведення випробувань отримано базу експериментальних даних (табл. 3).

Таблиця 3. Фрагмент бази експериментальних даних  $R$ , Ом і  $T$ ,  $^{\circ}\text{C}$

Номер вимірювання	$R_{\text{NTC } \#1}$	$T_{\text{Pt100}}$	$T_{\text{NTC } \#2}$	$T_{\text{NTC } \#3}$	$T_{\text{NTC } \#4}$	$T_{\text{NTC } \#5}$	$T_{\text{NTC } \#6}$
1	963087,30	-55,438	-55,239	-55,199	-55,171	-55,225	-55,150
2	959686,94	-55,314	-55,020	-54,986	-54,966	-55,017	-54,950
3	949599,80	-55,167	-54,868	-54,828	-54,805	-54,857	-54,787
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3704	185,34	149,810	149,867	149,621	149,454	149,791	149,772

Аналіз експериментальних даних (табл. 3) виконували за зведеною похибкою (5), оцінку якої отримали на основі відібраних з кроком  $5\text{ }^{\circ}\text{C}$   $n = 42$  значень температури п'яти термісторів TDK B57861S №2–№6 у діапазоні  $-55\dots150\text{ }^{\circ}\text{C}$ . Максимальні значення зведеної похибки  $\Delta_{\text{NTC}}^{\text{ЗВ}}$  термісторів TDK B57861S наведено в табл. 4.

**Таблиця 4.** Максимальні значення зведеної похибки термісторів TDK B57861S

NTC-термістор	№2	№3	№4	№5	№6
$\Delta_{NTC}^{ЗВ}, \%$	0,1390	0,0732	0,1590	0,0634	0,0937

Отриману базу даних (див. табл. 3) використовували для проведення обчислювальних експериментів на підставі навчальної вибірки RBF-мереж, у якій входом є електричний опір  $R_{NTC, \text{№}1}$ , а виходом — опорне значення температури  $T_{Pt100}$ .

### ОБЧИСЛЮВАЛЬНИЙ ЕКСПЕРИМЕНТ

Завданням обчислювального експерименту є створення RBF-мереж для моделювання ГХ NTC-термісторів (1) і оцінювання точності отриманих результатів.

Методика оброблення вимірювальної інформації з використанням RBF-мереж містить такі етапи:

1. Формування навчальної вибірки RBF-мереж.
2. Визначення кількості нейронів прихованого шару RBF-мереж.
3. Моделювання всіх можливих варіантів архітектур RBF-мереж з урахуванням максимальної кількості нейронів прихованого шару.
4. Навчання і збереження  $m$  кращих RBF-мереж відповідно до мінімальних значень критеріїв помилок навчання, контролю та тестування нейромережевих моделей.
5. Вибір з  $m$  нейромережевих моделей найбільш адекватної RBF-мережі на підставі мінімального значення критерію дисперсійного відношення.
6. Отримання інтерпольованого значення температури в робочому діапазоні NTC-термісторів.
7. Оцінювання достовірності та точності результатів нейромережевої апроксимації з використанням коефіцієнтів парної кореляції та статистичних критеріїв MAD, MSE, MAPE і MPE.
8. Повторення етапів 4–7 для перевірки збіжності та відтворюваності отриманих результатів нейромережевого моделювання ГХ NTC-термісторів під час повторних навчань RBF-мереж на різних апаратних платформах.

Оскільки одиницю вимірювання температури в залежності (1) взято за шкалою Кельвіна, то розмірність даних, отриманих в результаті фізичного експерименту (див. табл. 3) та використаних для створення навчальної вибірки RBF-мереж, наведено у кельвінах (K) (табл. 5).

**Таблиця 5.** Фрагмент навчальної вибірки RBF-мереж для термістора TDK B57861S №1

Номер вимірювання	$R_{NTC, \text{№}1}, \text{ Ом}$	$T_{Pt100}, \text{ К}$
1	963087,30	217,712
2	959686,94	217,836
3	949599,80	217,983
⋮	⋮	⋮
3704	185,34	422,960

У системі статистичного аналізу даних STATISTICA 6.1 з використанням інструменту «Мастер решений» модуля «Нейронные сети» визначали діапазон нейронів прихованого шару архітектури RBF-мереж (рис. 6).

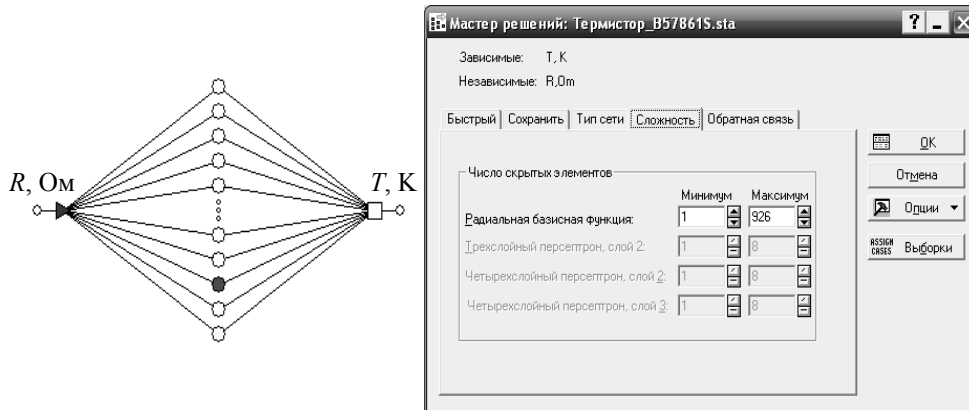


Рис. 6. Загальний вигляд архітектури RBF-мереж і скріншот діалогового вікна для вибору кількості нейронів прихованого шару

Під час моделювання всіх можливих варіантів архітектур НМ з RBF функціями активації (4) з урахуванням максимальної кількості нейронів прихованого шару 926 (рис. 6), яка залежить від обсягу навчальної вибірки, виконували 10 повторних навчань 100 нейромережевих моделей (10×100=1000). Для забезпечення збіжності та відтворюваності результатів нейромережевої апроксимації ГХ TDK B57861S №1 обчислювальний експеримент повторювали тричі на одній апаратній платформі та двічі на двох різних апаратних платформах. При цьому на першому етапі з кожних ста моделей НМ зберігали п'ять RBF-мереж за мінімальними значеннями критеріїв помилки навчання, контрольної помилки та тестової помилки (рис. 7, 8).

Подробные результаты моделей (Термистор_B57861S.sta)							
N	Архитектура	Производительность обуч.	Контр. производительность	Тест. производительность	Ошибка обучения	Контрольная ошибка	Тестовая ошибка
1	РБФ 1:1-68-1:1	0,000603	0,000645	0,000617	0,000010	0,000011	0,000011
2	РБФ 1:1-87-1:1	0,000578	0,000639	0,000596	0,000010	0,000011	0,000011
3	РБФ 1:1-101-1:1	0,000555	0,000629	0,000578	0,000010	0,000011	0,000010
4	РБФ 1:1-111-1:1	0,000467	0,000492	0,000494	0,000008	0,000009	0,000009
5	РБФ 1:1-134-1:1	0,000381	0,000424	0,000417	0,000007	0,000007	0,000008

Рис. 7. Результат навчання п'яти кращих RBF-мереж у модулі «Нейронные сети» системи STATISTICA 6.1 для першого обчислювального експерименту серії 1

Подробные результаты моделей (Термистор_B57861S.sta)							
N	Архитектура	Производительность обуч.	Контр. производительность	Тест. производительность	Ошибка обучения	Контрольная ошибка	Тестовая ошибка
1	РБФ 1:1-119-1:1	0,000789	0,000899	0,000962	0,000014	0,000016	0,000017
2	РБФ 1:1-141-1:1	0,000445	0,000493	0,000466	0,000008	0,000009	0,000008
3	РБФ 1:1-143-1:1	0,000379	0,000394	0,000405	0,000007	0,000007	0,000007
4	РБФ 1:1-215-1:1	0,000309	0,000355	0,000368	0,000005	0,000006	0,000007
5	РБФ 1:1-236-1:1	0,000317	0,000352	0,000401	0,000005	0,000006	0,000007

Рис. 8. Результат навчання п'яти кращих RBF-мереж у модулі «Нейронные сети» системи STATISTICA 6.1 для другого обчислювального експерименту серії 1

З отриманих п'яти моделей RBF-мереж обирали одну за мінімальним значенням статистичного критерію дисперсійного відношення, який регла-

ментується нормативною документацією для перевірки адекватності побудованої моделі ГХ експериментальним даним. Так, наприклад, за значенням цього критерію для першого обчислювального експерименту обрано модель RBF-мережі з архітектурою 1-134-1, а для другого — модель RBF-мережі з архітектурою 1-215-1. У результаті п'яти серій обчислювальних експериментів із 5000 створених моделей RBF-мереж відібрано 50 моделей, архітектури яких наведено в табл. 6 [26].

**Таблиця 6.** Відібрані моделі RBF-мереж для апроксимації ГХ термістора TDK B57861S №1

Номер експерименту	Платформа				
	AMD Athlon II X2 240, 2,81 ГГц, 1,75 ГБ RAM			AMD Turion II Dual-Core Mobile M520, 779 МГц, 3,00 ГБ RAM	Intel Celeron CPU G530, 2,40 ГГц, 1,91 ГБ RAM
	Серія 1	Серія 2	Серія 3	Серія 4	Серія 5
1	1-134-1	1-137-1	1-124-1	1-181-1	1-116-1
2	1-215-1	1-198-1	1-229-1	1-169-1	1-143-1
3	1-102-1	1-172-1	1-167-1	1-126-1	1-152-1
4	1-180-1	1-113-1	1-179-1	1-174-1	1-172-1
5	1-127-1	1-157-1	1-119-1	1-145-1	1-102-1
6	1-174-1	1-178-1	1-128-1	1-134-1	1-156-1
7	1-119-1	1-156-1	1-138-1	1-201-1	1-137-1
8	1-236-1	1-180-1	1-97-1	1-137-1	1-216-1
9	1-172-1	1-127-1	1-216-1	1-182-1	1-128-1
10	1-187-1	1-151-1	1-203-1	1-99-1	1-201-1

## РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНОГО ЕКСПЕРИМЕНТУ

Для кожної серії експериментів з використанням відібраних моделей RBF-мереж визначали інтерпольовані значення температури (табл. 7–11).

**Таблиця 7.** Фрагмент інтерпольованих значень температури  $T$ , К для першої серії експерименту

Номер вимірювання	Вхід $R$ , Ом	Архітектура									
		1-134-1	1-215-1	1-102-1	1-180-1	1-127-1	1-174-1	1-119-1	1-236-1	1-172-1	1-187-1
1	963087,30	217,825	217,726	217,790	217,696	217,746	217,791	217,758	217,740	217,772	217,707
2	959686,94	217,865	217,765	217,837	217,765	217,798	217,838	217,802	217,796	217,819	217,777
3	949599,80	217,992	217,894	217,980	217,963	217,952	217,980	217,942	217,960	217,964	217,964
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3704	185,34	422,987	422,969	422,928	423,084	422,970	423,068	422,981	422,961	422,985	422,965

**Таблиця 8.** Фрагмент інтерпольованих значень температури  $T$ , К для другої серії експерименту

Номер вимірювання	Вхід $R$ , Ом	Архітектура									
		1-137-1	1-198-1	1-172-1	1-113-1	1-157-1	1-178-1	1-156-1	1-180-1	1-127-1	1-151-1
1	963087,30	217,731	217,811	217,809	217,761	217,795	217,751	217,851	217,794	217,831	217,793
2	959686,94	217,786	217,853	217,852	217,813	217,843	217,804	217,883	217,814	217,867	217,839
3	949599,80	217,951	217,984	217,984	217,966	217,987	217,965	217,992	217,905	217,984	217,979
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3704	185,34	423,086	423,005	422,998	422,909	422,984	422,944	423,081	422,880	423,039	422,945

**Таблиця 9.** Фрагмент інтерпольованих значень температури  $T$ , К для третьої серії експерименту

Номер вимірювання	Вхід $R$ , Ом	Архітектура									
		1-124-1	1-229-1	1-167-1	1-179-1	1-119-1	1-128-1	1-138-1	1-97-1	1-216-1	1-203-1
1	963087,30	217,728	217,799	217,797	217,777	217,788	217,789	217,787	217,819	217,719	217,817
2	959686,94	217,787	217,843	217,841	217,815	217,836	217,809	217,832	217,833	217,788	217,854
3	949599,80	217,957	217,978	217,977	217,942	217,981	217,897	217,971	217,899	217,969	217,975
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3704	185,34	422,996	423,034	422,967	422,965	422,945	422,982	423,032	422,921	422,994	422,989

**Таблиця 10.** Фрагмент інтерпольованих значень температури  $T$ , К для четвертої серії експерименту

Номер вимірювання	Вхід $R$ , Ом	Архітектура									
		1-181-1	1-169-1	1-126-1	1-174-1	1-145-1	1-134-1	1-201-1	1-137-1	1-182-1	1-99-1
1	963087,30	217,736	217,712	217,845	217,812	217,801	217,756	217,764	217,834	217,735	217,786
2	959686,94	217,790	217,784	217,876	217,842	217,843	217,808	217,811	217,873	217,786	217,821
3	949599,80	217,953	217,968	217,988	217,948	217,974	217,963	217,956	217,995	217,944	217,932
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3704	185,34	423,018	422,951	423,109	422,984	423,064	422,891	422,958	423,075	422,958	422,989



**Таблиця 11.** Фрагмент інтерпольованих значень температури  $T$ , °К для п'ятої серії експерименту

Номер вимірювання	Вхід $R$ , Ом	Архітектура									
		1-116-1	1-143-1	1-152-1	1-172-1	1-102-1	1-156-1	1-137-1	1-216-1	1-128-1	1-201-1
1	963087,30	217,728	217,819	217,748	217,787	217,881	217,793	217,857	217,680	217,843	217,737
2	959686,94	217,778	217,861	217,800	217,836	217,912	217,836	217,889	217,758	217,870	217,795
3	949599,80	217,930	217,991	217,954	217,982	218,012	217,974	217,997	217,966	217,971	217,970
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3704	185,34	422,921	422,853	422,979	422,965	422,950	423,027	423,004	423,247	423,016	423,217

### ОБГОВОРЕННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

Достовірність отриманих результатів нейромережевого моделювання підтверджується високим значенням коефіцієнтів парної кореляції (рис. 9–13) між опорними значеннями температури сенсора Pt100 (табл. 5) та значеннями температури, інтерпольованими на основі RBF-мереж (див. табл. 7–11).

Переменная	Т, К
1-134-1	0,999999919638
1-215-1	0,999999943495
1-102-1	0,999999906847
1-180-1	0,999999934319
1-127-1	0,999999914442
1-174-1	0,999999944250
1-119-1	0,999999915800
1-236-1	0,999999954660
1-172-1	0,999999932711
1-187-1	0,999999940731

Рис. 9. Скріншот розрахунку коефіцієнтів парної кореляції для першої серії експериментів

Переменная	Т, К
1-137-1	0,999999925479
1-198-1	0,999999947078
1-172-1	0,999999938889
1-113-1	0,999999908431
1-157-1	0,999999926207
1-178-1	0,999999939497
1-156-1	0,999999932327
1-180-1	0,999999943140
1-127-1	0,999999924697
1-151-1	0,999999930302

Рис. 10. Скріншот розрахунку коефіцієнтів парної кореляції для другої серії експериментів

Корреляции (Серия 3.sta)	
Отмеченные корреляции значимы на уровне $p < ,00001$	
N=3704 (Построчное удаление ПД)	
Переменная	Т, К
1-124-1	0,999999926261
1-229-1	0,999999938357
1-167-1	0,999999937264
1-179-1	0,999999936980
1-119-1	0,999999911144
1-128-1	0,999999919835
1-138-1	0,999999927132
1-97-1	0,999999886704
1-216-1	0,999999945933
1-203-1	0,999999945451

Рис. 11. Скріншот розрахунку коефіцієнтів парної кореляції для третьої серії експериментів

Корреляции (Серия 4.sta)	
Отмеченные корреляции значимы на уровне $p < ,00001$	
N=3704 (Построчное удаление ПД)	
Переменная	Т, К
1-181-1	0,999999937955
1-169-1	0,999999935719
1-126-1	0,999999910426
1-174-1	0,999999936261
1-145-1	0,999999927157
1-134-1	0,999999914954
1-201-1	0,999999937330
1-137-1	0,999999920399
1-182-1	0,999999940962
1-99-1	0,999999892853

Рис. 12. Скріншот розрахунку коефіцієнтів парної кореляції для четвертої серії експериментів

Корреляции (Серия 5.sta)	
Отмеченные корреляции значимы на уровне $p < ,00001$	
N=3704 (Построчное удаление ПД)	
Переменная	Т, К
1-116-1	0,999999919356
1-143-1	0,999999919706
1-152-1	0,999999932952
1-172-1	0,999999945724
1-102-1	0,999999905159
1-156-1	0,999999930199
1-137-1	0,999999924071
1-216-1	0,999999939316
1-128-1	0,999999921568
1-201-1	0,999999929571

Рис. 13. Скріншот розрахунку коефіцієнтів парної кореляції для п'ятої серії експериментів

Оцінку точності отриманих результатів нейромережевого моделювання виконували в системі STATISTICA 6.1 з використанням критеріїв MAD

(Mean Absolute Derivation), MSE (Mean Squared Error), MAPE (Mean Absolute Percentage Error) і MPE (Mean Percentage Error):

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^n |T_{\text{Pt100}}^i - \hat{T}_{\text{RBF}}^i|; \quad (6)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (T_{\text{Pt100}}^i - \hat{T}_{\text{RBF}}^i)^2; \quad (7)$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \frac{|T_{\text{Pt100}}^i - \hat{T}_{\text{RBF}}^i|}{T_{\text{Pt100}}^i}; \quad (8)$$

$$\text{MPE} = \frac{100}{n} \sum_{i=1}^n \frac{(T_{\text{Pt100}}^i - \hat{T}_{\text{RBF}}^i)}{T_{\text{Pt100}}^i}, \quad (9)$$

де  $T_{\text{Pt100}}^i$  — опорні значення температури сенсора Pt100, які отримано внаслідок фізичного експерименту (див. табл. 5);  $\hat{T}_{\text{RBF}}^i$  — інтерпольовані значення температури з використанням моделей RBF-мереж.

Мінімальні та максимальні значення критеріїв (6)–(9) для п'яти серій обчислювальних експериментів наведено в табл. 12 [25, 26].

**Таблиця 12.** Діапазони значень критеріїв MAD, MSE, MAPE, MPE для п'яти серій обчислювальних експериментів

Номер серії	MAD <sub>min</sub>	MAD <sub>max</sub>	MSE <sub>min</sub>	MSE <sub>max</sub>	MAPE <sub>min</sub>	MAPE <sub>max</sub>	MPE <sub>min</sub>	MPE <sub>max</sub>
1	0,0131	0,0192	0,00032	0,00065	0,00399	0,00583	-0,00021	0,00054
2	0,0145	0,0188	0,00037	0,00064	0,00441	0,00567	-0,00029	0,00017
3	0,0141	0,0212	0,00037	0,00064	0,00428	0,00649	-0,00010	0,00067
4	0,0152	0,0202	0,00041	0,00075	0,00464	0,00619	-0,00007	0,00021
5	0,0147	0,0193	0,00038	0,00066	0,00451	0,00589	-0,00012	0,00016

Перетин інтервалів (табл. 12) і порядок оцінок статистичних критеріїв (6)–(9) свідчать про збіжність і відтворюваність отриманих результатів обчислювальних експериментів, що характеризує високу точність нейромережевої апроксимації ГХ NTC-термістора TDK B57861S №1 у робочому діапазоні температур. Отримані результати нейромережевого моделювання дозволяють рекомендувати використання RBF-мереж для автоматизації процедури побудови ГХ при періодичному калібруванні термісторів на стадії експлуатації за умови, якщо зведена похибка температури  $\Delta_{\text{RBF}}^{\text{ЗБ}}$  найменш точних RBF-мереж, які створені в кожній серії обчислювальних експериментів, буде меншою за зведену похибку  $\Delta_{\text{NTC}}^{\text{ЗБ}}$  NTC-термісторів №2–№6 (див. табл. 4).

Для розрахунку зведеної похибки з бази даних результатів обчислювальних експериментів у діапазоні  $-55 \dots 150$  °C з кроком 5 °C відібрано  $n=42$  значення температури найменш точних за критерієм (8) RBF-мереж (табл. 13).

Таблиця 13. Відібрані значення  $T$ , °C сенсора Pt100 і RBF-мереж

$T_{Pt100}$	$T_{RBF\ 1-102-1}$	$T_{RBF\ 1-113-1}$	$T_{RBF\ 1-97-1}$	$T_{RBF\ 1-99-1}$	$T_{RBF\ 1-101-1}$
-55,029	-55,027	-55,033	-55,154	-55,100	-55,028
-50,142	-50,147	-50,146	-50,113	-50,147	-50,137
-45,110	-45,110	-45,117	-45,100	-45,106	-45,099
-40,033	-40,067	-40,068	-40,073	-40,074	-40,056
-35,031	-35,028	-35,026	-35,031	-35,034	-35,035
-30,174	-30,182	-30,176	-30,176	-30,176	-30,182
-25,073	-25,104	-25,093	-25,089	-25,098	-25,099
-20,005	-20,009	-20,007	-20,012	-20,015	-20,011
-15,087	-15,099	-15,112	-15,097	-15,117	-15,095
-10,045	-10,091	-10,088	-10,088	-10,103	-10,090
-5,044	-5,008	-5,007	-5,015	-5,020	-5,010
0,006	0,020	0,009	0,011	0,011	0,014
5,041	5,008	5,001	4,999	5,003	4,995
10,027	10,034	10,030	10,033	10,038	10,028
15,181	15,142	15,135	15,144	15,142	15,138
20,049	20,046	20,048	20,052	20,045	20,046
25,033	25,029	25,027	25,015	25,024	25,005
30,038	30,022	30,024	30,020	30,021	30,025
35,041	35,040	35,032	35,029	35,026	35,029
40,100	40,116	40,105	40,097	40,100	40,097
45,024	45,034	45,019	45,021	45,028	45,019
50,008	50,014	49,998	49,993	50,017	50,002
55,002	54,993	54,984	54,985	54,999	54,996
60,020	60,010	60,004	60,013	60,019	60,016
65,071	65,056	65,063	65,048	65,048	65,047
70,063	70,019	70,041	70,034	70,030	70,025
75,032	75,010	75,002	74,998	75,007	74,993
80,069	80,037	80,036	80,048	80,040	80,055
85,034	85,024	85,020	85,018	85,014	85,018
90,056	90,027	90,011	90,015	90,020	90,015
95,048	95,033	95,032	95,030	95,035	95,022
100,051	100,018	99,999	99,998	100,031	99,999
105,043	105,005	104,990	104,990	105,039	104,995
110,073	110,006	110,030	110,043	110,063	110,038
115,027	114,995	114,960	114,947	114,961	114,974
120,085	120,127	120,117	120,081	120,076	120,104
125,057	125,051	125,069	125,045	125,032	125,035
130,058	130,008	130,006	130,025	130,027	129,996
135,001	134,916	134,944	134,949	134,985	134,964
140,084	140,057	140,063	140,056	140,052	140,055
145,129	145,091	145,079	145,069	145,108	145,059
149,810	149,778	149,759	149,771	149,839	149,800

Зведену похибку  $\Delta_{RBF}^{3B}$  розраховували за формулою

$$\Delta_{RBF}^{3B} = \left( \frac{\hat{T}_{RBF} - T_{Pt100}}{T_p} \right) \cdot 100, \quad (10)$$

де  $\hat{T}_{RBF}$  і  $T_{Pt100}$  — відповідно значення температури, отримані з використанням RBF-мереж, і опорні значення температури;  $T_p$  — протяжність робочого діапазону температури термісторів.

Максимальні значення  $\Delta_{RBF}^{3B}$  моделей RBF-мереж (10) наведено в табл. 14.

**Таблиця 14.** Максимальні значення зведеної похибки найменш точних моделей RBF-мереж

Архітектура	1-102-1	1-113-1	1-97-1	1-99-1	1-101-1
$\Delta_{RBF}^{3B}$ , %	0,0205	0,0181	0,0142	0,0141	0,0166

Аналіз результатів фізичного (табл. 4) і обчислювального (табл. 14) експериментів показує, що у робочому діапазоні температур будь-яке максимальне значення  $\Delta_{RBF}^{3B}$  найменш точних RBF-мереж у кілька разів менше за будь-яке максимальне значення зведеної похибки вимірювання температури  $\Delta_{NTC}^{3B}$  з використанням термісторів TDK B57861S №2–№6.

Результат оцінки значущості відмінностей  $\Delta_{NTC}^{3B}$  і  $\Delta_{RBF}^{3B}$ , отриманий на основі *t*-тесту для незалежних змінних у системі STATISTICA 6.1, свідчить про відмінність похибок температури п'яти сенсорів і п'яти моделей RBF-мереж на рівні  $p=0,0014$ , що підтверджує високу точність нейромережевої апроксимації ГХ NTC-термісторів (рис. 14, 15).

Т-критерий независимых выборок (Погрешности_TDK_sta)					
Замечание: Переменные рассм. как независимые выборки					
Группа 1 и Группа 2	Среднее Группа 1	Среднее Группа 2	t-знач.	ст. св.	p
$\Delta_{NTC}^{3B}$ vs. $\Delta_{RBF}^{3B}$	0,105660	0,016700	4,765768	8	0,001416

Рис. 14. Скріншот результату оцінки значущості відмінностей зведених похибок  $\Delta_{NTC}^{3B}$  і  $\Delta_{RBF}^{3B}$

Оцінку похибки нейромережевої апроксимації  $\Delta_{RBF}$  ГХ NTC-термістора TDK B57861S №1 визначали з використанням критерію MAD (6) і відібраних  $n=42$  значень температури найменш точних RBF-мереж (табл. 15) [25].

Порядок оцінок  $\Delta_{RBF}$  (табл. 15) свідчить про те, що похибка нейромережевої апроксимації ГХ TDK B57861S №1 не менше ніж в 1,5 разу нижча за похибку математичного перетворення  $\Delta = 0,05$  °C залежності (1), яка використовується в обчислювальному алгоритмі СЗД Agilent 34970A.



Рис. 15. Діаграма розкиду значень зведених похибок  $\Delta_{NTC}^{ЗВ}$  і  $\Delta_{RBF}^{ЗВ}$

**Таблиця 15.** Похибка найменш точних моделей RBF-мереж

Архітектура	1-102-1	1-113-1	1-97-1	1-99-1	1-101-1
$\Delta_{RBF}, ^\circ\text{C}$	0,0230	0,0242	0,0267	0,0207	0,0235

Таким чином, аналіз отриманих результатів дозволяє рекомендувати застосування RBF-мереж в обчислювальних алгоритмах оброблення вимірювальної інформації сучасних СЗД для підвищення точності градуювання і автоматизації процедури періодичного калібрування NTC-термісторів.

## ВИСНОВКИ

Розроблено метод нейромережевого оброблення вимірювальної інформації для апроксимації градуювальних характеристик NTC-термісторів у робочому діапазоні температур, застосування якого дозволяє підвищити точність сучасних СЗД, які використовуються у випробувальних і калібрувальних лабораторіях, за рахунок застосування моделей RBF-мереж в обчислювальних алгоритмах оброблення вимірювальної інформації.

Встановлено, що застосування створеного методу дозволяє отримати максимальну похибку перетворення опору в температуру  $\Delta_{RBF}=0,027\text{ }^\circ\text{C}$  і не менше ніж в 1,5 рази знизити похибку апроксимації градуювальної характеристики NTC-термісторів порівняно з допустимою похибкою математичного перетворення  $\Delta=0,05\text{ }^\circ\text{C}$  поліноміальної моделі третього порядку, яка використовується у програмному забезпеченні сучасних систем збирання та оброблення вимірювальної інформації.

## ЛІТЕРАТУРА

1. *Metrology in industry: the key for quality*; edited by French College of Metrology. Paris, 2006, 270 p.

2. E.S. Polishchuk, V.M. Vanko, M.M. Dorozhovets, V.O. Yatsuk, and Yu.V. Yatsuk, *Measuring transducers (sensors): textbook*. Lviv: National Lviv Polytechnic University, 2015, 584 p.
3. V.M. Sharapov et al., *Sensors: ref. allowance*. M.: Technosphere, 2012, 624 p.
4. R.I. Baitsar and S.S. Warsaw, *Semiconductor microsensors: academic. manual*. Lviv: National Lviv Polytechnic University, 2001, 290 p.
5. Y.T. Lutsik, O.P. Guk, O.I. Lakh, and B.I. Stadnyk, *Temperature measurement. Theory and practice*. Lviv: Beskyd Bit, 2006, 580 p.
6. John S. Steinhart and Stanley R. Hart, "Calibration curves for thermistors," *Deep Sea Research and Oceanographic Abstracts*, vol. 15, issue 4, August 1968, pp. 497–503.
7. *Agilent 34970 A Data Acquisition. Switch Unit User Guide Edition 3.0*. Agilent Technologies, Inc., 2003, 434 p.
8. S. Khaikin, *Neural networks: a complete course*; ed. 2nd. M.: Williams, 2006, 1104 p.
9. O.A. Gerashchenko et al., *Temperature measurements: a reference book*. K.: Naukova dumka, 1989, 704 p.
10. V.A. Granovsky and T.N. Siraya, *Methods of processing experimental data during measurements*. L: Energoatomizdat, 1990, 288 p.
11. R.G. Jackson, *The latest sensors*. M: Technosfera, 2007, 384 p.
12. V.T. Kondratov, *Fundamentals of the theory of automatic correction of systematic changes in the correction of physical quantities with an unstable and non-linear function of the sensor transformation: abstract of the thesis. thesis. Dr. tech. Sciences: 05.11.15 and 05.11.01*. Kyiv, 2001, 36 p.
13. J. Frieden, *Modern sensors: a reference book*. M: Technosfera, 2005, 592 p.
14. O.V. Zaporozhets, V.A. Korotenko, and T.A. Ovcharova, "Compensation for the nonlinearity of the conversion function of measuring devices using an artificial neural network," *Management, Navigation and Communication Systems*, no. 4(16), pp. 99–103, 2010.
15. A.V. Degtyarev, O.V. Zaporozhets, and T.A. Ovcharova, "Adaptive system for compensating the nonlinearity of the conversion function of measuring devices based on a three-layer perceptron," *Electrical And Computer Systems*, no. 6, pp. 235–241, 2012.
16. O.V. Zaporozhets, T.A. Ovcharova, and I.V. Ruzhentsev, "Compensation for nonlinearity of a semiconductor thermistor using an artificial neural network," *Information Processing Systems*, no. 6(131), pp. 64–67, 2015.
17. Nicholas J. Cotton and Bogdan M. Wilamowski, "Compensation of Sensors Nonlinearity with Neural Networks," *24th IEEE International Conference on Advanced Information Networking and Applications, 2010*, pp. 1210–1216.
18. T.D. Dongale and R.K. Kamat, "Modelling of NTC-thermistor using artificial neural network for nonlinearity compensation," *Informatcs Engineering, an International Journal (IEIJ)*, 2013, vol. 1, pp. 15–20.
19. Shakeb A. Khan, Tarikul Islam, and Gulshan, "Artificial Neural Network based online sensor calibration and compensation," *Computing*, vol. 6, no. 3, pp. 74–78, 2007.
20. S.S. Fedin and I.S. Zubretska, "Ensuring the accuracy of approximation of the R/T-characteristic of an NTC-thermistor based on neural network modeling," *Technology and Design in Electronic Equipment*, no. 4, pp. 28–35, 2015.
21. N.A. Zubretska, S.S. Fedin, and I.S. Zubretska, "Construction of calibration characteristics of temperature sensors using neural networks," *Bulletin of the Engineering Academy of Ukraine*, no. 1, pp. 130–136, 2015.
22. M.J.D. Powell, "The theory of radial basis function approximation in 1990," in W. Light, ed., *Advances in Numerical Analysis Vol. II: Wavelets, Subdivision Algorithms and Radial Basis Functions*. Oxford: Oxford Science Publications, 1992, pp. 105–210.

23. *Information Technology. Glossary of terms: DSTU ISO/IEC 2382:2017*. K.: SE “UkrNDNC”, 2020, 464 p.
24. S. Osovsky, *Neural networks for information processing*. M.: Finance and statistics, 2004, 344 p.
25. S.S. Fedin and I.S. Zubretska, *Neural network approximation of calibration characteristics of NTC-thermistors: monograph*. K.: Interservice, 2017, 196 p.
26. S.S. Fedin, I.S. Zubretska, and O.O. Polikarpov, “Ensuring the accuracy of construction of calibration characteristics of NTC-thermistors based on neural networks with radial basis functions,” *Metrology and Devices*, no. 1(63), pp. 37–46, 2017.

Received 25.08.2022

#### INFORMATION ON THE ARTICLE

**Serhii S. Fedin**, ORCID: 0000-0001-9732-632X, National Transport University, Ukraine, e-mail: sergey.fedin1975@gmail.com

**Irina S. Zubretska**, ORCID: 0000-0002-6583-3690, “Trade City”, Ukraine, e-mail: pfar-seer@gmail.com

**METHODS AND MODELS OF NEURAL NETWORKS FOR APPROXIMATION OF CALIBRATION CHARACTERISTICS OF NTC-THERMISTORS** / S.S. Fedin, I.S. Zubretska

**Abstract.** The hypothesis about the expediency of using RBF-networks to improve the accuracy of constructing the calibration characteristics of NTC-thermistors in the operating temperature range without dividing it into subranges is confirmed. It has been established that the error of the neural network approximation of the calibration characteristics of NTC-thermistors based on RBF-networks is at least one and a half times less than the permissible error of approximation of the third-order polynomial model, which is used in the software of modern systems for collecting and processing measurement information. A technique has been developed for processing measurement information using adaptive RBF-networks to automate constructing individual calibration characteristics and periodic calibration of NTC-thermistors.

**Keywords:** accuracy, measuring data, calibration, NTC-thermistor, operating temperature range, transformation function, neural network approximation, RBF-network.



## COST EFFECTIVE HYBRID GENETIC ALGORITHM FOR WORKFLOW SCHEDULING IN CLOUD

SANDEEP KUMAR BOTHRA, SUNITA SINGHAL, HEMLATA GOYAL

**Abstract.** Cloud computing plays a significant role in everyone's lifestyle by snugly linking communities, information, and trades across the globe. Due to its NP-hard nature, recognizing the optimal solution for workflow scheduling in the cloud is a challenging area. We proposed a hybrid meta-heuristic cost-effective load-balanced approach to schedule workflow in a heterogeneous environment. Our model is based on a genetic algorithm integrated with predict earliest finish time (PEFT) to minimize makespan. Instead of assigning the task randomly to a virtual machine, we apply a greedy strategy that assigns the task to the lowest-loaded virtual machine. After completing the mutation operation, we verify the dependency constraint instead of each crossover operation, which yields a better outcome. The proposed model incorporates the virtual machine's performance variance as well as acquisition delay, which concedes the minimum makespan and computing cost. One of the most astounding aspects of our cost-effective hybrid genetic algorithm (CHGA) is its capacity to anticipate by creating an optimistic cost table (OCT) while maintaining quadratic time complexity. Based on the results of our meticulous experiments on some real-world workflow benchmarks and comprehensive analysis of some recently successful scheduling algorithms, we concluded that the performance of our CHGA is melodious. CHGA is 14.58188%, 11.40224%, 11.75306%, and 9.78841% cheaper than standard Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Cost Effective Genetic Algorithm (CEGA), and Cost-Effective Load-balanced Genetic Algorithm (CLGA), respectively.

**Keywords:** cloud computing, cost effective, genetic algorithm, metaheuristic algorithm, predict earliest finish time, Workflow scheduling.

### INTRODUCTION

Cloud computing is a buzzword in the current era, which provides a very elastic 'pay as you go' model [1]. Dr. Raj Kumar Buyya says, "A cloud is a kind of parallel and distributed system made up of a number of linked, virtualized computers that are constantly provided and shown as one or more unified computing resources in accordance with service-level agreements negotiated between the service provider and customers" [2]. On the basis of physical location and distribution, various deployment models are available now. Task scheduling is critical for maximizing the use of cloud resources as well as providing end users with quality of service (QoS) [3].

Static scheduling and dynamic scheduling are two different sorts of task scheduling issues. In the static category, all task characteristics, including the costs of computation and communication for each activity as well as how those activities relate to one another, are known in advance. However, the dynamic category makes such information unavailable and makes judgments at runtime [4]. Furthermore, static scheduling refers to compile-time scheduling, and dynamic scheduling refers to scheduling at runtime.

Heuristic-based and guided random search-based algorithms are the two types of static scheduling algorithms that are most frequently used.

Heuristic-based algorithms deliver approximate, frequently excellent results because of their polynomial time complexity [5]. Similar to heuristic-based algorithms, guided random search-based algorithms provide approximations, but the results' quality may be increased by including more rounds, which raises the cost of the methods [6].

Dependent tasks are represented as a workflow, which is a set of nodes and edges, with each node representing a job and each edge representing follow-up dependence [7]. Workflow is scheduled and executed by the workflow management system where tasks are scheduled and provisioned to virtual machines [8]. Various researchers are engaged themselves to resolve the problem of resource scheduling in cloud. Many tasks have been completed using the heuristic approach; however it is not very excellent owing to its problem-dependent aspect, which is that it is unable to provide a globally optimum solution. As a result, the researcher prefers to use a meta-heuristic technique. Due to its task-independent character, the meta-heuristic method delivers a global optimal solution. The researcher's ultimate objective is to maximize cloud resource usage while lowering costs for cloud's end users [9], [10].

The majority of quadratic time complexity list-based scheduling algorithms just evaluate the current task when allocating a task to a processor. Although it is a low-cost method, it does not examine what comes before the current job, which could lead to poor decisions in some cases. Lookahead [11] is an example of an algorithm that analyses the impact on child nodes, but it raises the time complexity to the fourth order. As a result, we used the PEFT approach in our proposed model "cost-effective hybrid genetic algorithm" (CHGA). One of PEFT's most amazing features is its ability to predict by making an OCT with optimistic costs while preserving quadratic time complexity.

**Motivation.** Following a comprehensive review, we motivated to resolve a research gap where many parameters, such as virtual machine (VM) performance variation, booting time, and shutdown time, as well as load balancing across VMs and minimize execution time in parallel using heuristics approaches, are not effectively addressed.

**Objective.** The goal of this research is to arrange the tasks of workflow in such a way that it reduces not only computation costs but also processing time while maintaining load balance among virtual machines. Our objective was to create a hybrid meta-heuristic technique for reducing processing time and expense while maintaining load balance across virtual machines under time constraint. During population initialization in genetic algorithm, we employed predict earliest finish time (PEFT) approach, which is significant for decreasing the makespan. We also took into account the time it takes for a virtual machine (VM) to boot up and performance fluctuations, both of which have an effect on computation time and execution cost. This represents the novelty of our model. To keep the load balanced among the virtual machines, we employed a greedy method [10] in our proposed model "cost-effective hybrid genetic algorithm" (CHGA).

The remaining sections of the paper are structured as follows. The second section goes through some background information. Sections III and IV contain problem definitions and details of our proposed model, respectively. The performance review may be found in Section V. Section V consists of two sub-sections: result analysis and discussion. Finally, Section VI draws the paper toward its conclusion.

## **RELATED WORK**

Comprehensive work has been done by us on various meta-heuristic algorithms in literature [9], some of them are genetic algorithm (GA) [12], ant colony optimization (ACO) [13], particle swarm optimization (PSO) [14], artificial bee colony algorithm (ABC) [15] etc.

The RDPSO (Revised Discrete PSO) technique employed in [16] involves a greedy adaptive search procedure to establish the swarm particle, followed by the computation of local best and global best. It focuses on achieving the lowest execution cost, but load balancing across virtual machines is not provided.

In literature [17], researchers designed a PSO-based algorithm to minimize execution cost as well as makespan and compared it with the Best Resource Selection (BRS) algorithm, but they didn't take into account dependent tasks in scheduling approach. This deficiency is removed by [18], where ACO is applied to the workflow. They used an approach ant strategy: front ant and back ant. Their study took into account pre-execution time and a pheromone threshold value, but they did not mimic a different type of scientific workflow.

Researchers in Dynamic Objective-based GA (DOGA) [19] reduced the cost of workflow execution and reached a result that was comparable to PSO, but they ignored the booting time factor and the load balancing approach. Authors provided a GA-based technique in the literature [20], where cost and time span are both reduced within a user-defined deadline. This paper was not based on real world workflow, which is accomplished by [21].

A multi-objective PSO approach with a weighted linear transform fitness function is presented in the literature [22] and they conclude that their proposed algorithm is better than genetic algorithms, but they consider only makespan and resource utilization as parameters, not other parameters like execution cost, load distribution, etc. The outcome of their experiment is not very trustworthy due to the limited size of their workflow.

A new approach SACO Slave ACO(SACO) [23] proposed a slave-ant concept where two techniques are used: diversification and reinforcement. These techniques escape slave ants from long paths. Their experiment didn't consider heterogeneous resources or load balance concepts. Multi Objectives ACO(MO-ACO) [24] addresses this flaw by presenting an approach for scheduling jobs in a cloud context that considers load balancing with cost and time but ignores dependent tasks in the cloud.

The Greedy-Ant-based ACO [25] approach uses forward and backward dependency techniques to build transition probability. To allocate the virtual machine, they used a greedy strategy. They compared their meta-heuristic model with a heuristic that has a high level of scarceness in their research.

In the suggested GA [26], VMs are grouped based on their capacity to shorten the time it takes for a procedure to complete. Before clustering the VM, they considered cost computing to make this approach more successful. They did not include the VM termination delay in their study, and they also did not examine the load balancing idea.

In the literature [27], authors focused on function optimization using improved genetic algorithms, whereas machine learning concepts are included with GA [28].

In order to reduce makespan and cost, authors presented a HEFT-ACO technique [29] that is based on the heterogeneous earliest end time (HEFT) and ACO, but they did not integrate the idea of load balancing across virtual machines.

In research [10], authors focused on balancing the load among virtual machines to increase performance. To achieve this, a greedy seeding strategy was applied with the genetic algorithm, but there was no efficient heuristic approach to reduce the makespan and cost.

Following a comprehensive review, we observed a research gap where many parameters, such as virtual machine (VM) performance variation, booting time, and shutdown time, as well as load balancing across VMs and minimize execution time in parallel using heuristics, are not effectively addressed.

Our goal was to develop a hybrid meta-heuristic approach for processing time and cost reduction in a time-constrained situation while maintaining load balance across virtual machines. To accomplish this, we used the PEFT strategy during population initialization, which helps to reduce the makespan. The ability of PEFT to anticipate by building an optimistic cost table (OCT) while preserving quadratic time complexity is one of its most amazing features. We also took into account the time it takes for a VM to boot up and performance fluctuations, both of which have an effect on computation time and execution cost. To keep the load balanced among the virtual machines, we employed a greedy method in our proposed model CHGA.

**PROBLEM DEFINITION**

Minimization of computing costs and makespan of scientific workflow with balancing the loads among virtual machines is the main motto of our proposed Cost Effective Hybrid Genetic Algorithm (CHGA), which works under a user-defined deadline constraint. A simple workflow is depicted in Fig. 1, and its corresponding encoding is represented in Fig. 2.

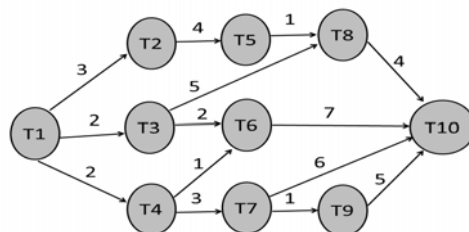


Fig. 1. Example of Workflow

	Encoding of Chromosome									
OrderOfTask	1	2	3	4	5	6	7	8	9	10
Task	T1	T4	T3	T6	T7	T2	T9	T5	T8	T10
VM	1	3	3	4	2	2	5	4	1	5
VmType	3	2	2	2	1	1	1	2	3	1

Fig. 2. Encoding of workflow depicted in Fig. 1

In a heterogeneous cloud computing environment, variation in the performance of VMs and booting time delays are two main factors that impact the makespan of the scientific workflow. That’s why we considered both of the above-mentioned parameters in our proposed model. Schedule is illustrated as

$S = \{VM_{SET}, Map, TET, TEC\}$ , Where  $VM_{SET}$  a virtual machine pool, and Map denotes the selection of an appropriate virtual machine to perform a task. Total Execution Time and Total Execution Cost, respectively, are abbreviated as TET and TEC. We generate the value 0%–24% randomly as a performance variation, and the acquisition delay is assumed to be 1 minute for each VM. We defined the problems to achieve our objectives. If TET violates the deadline constraint, then TEC is not computed, otherwise it will be computed.

## THE PROPOSED HYBRID GENETIC ALGORITHM

### Description of the CHGA

We explained CHGA step-by-step here.

**Step 1.** During population initialization, the chromosome is encoded in the same way as in the meta-heuristic technique provided by [25]. OrderOfTask, Task, VM, and VmType are four fields that are used to encode a chromosome.

If the total population is  $N$ , then  $(N - 1)$  is initialized using a random technique and the remaining is using PEFT. PEFT is described in section IV B.

**Step 2.** During the population initialization, we employed a greedy technique [10] to balance the load among several virtual machines, as illustrated through flowchart in this paper. This strategy assigns the new task to those VMi which have minimum load at that time. Compute Load  $L_i$  on a VMi:

$$L_i = \frac{\sum_{j=1}^n T_j}{VMC_i} .$$

**Step 3.** Now compute the fitness of each candidate.

Step 3.1: Calculate the execution and transfer times for all of the individual's tasks.

Divide the task's size  $Size_{T_i}$  by the virtual machine's processing speed  $Speed_{VM_k}$  to find the task's execution time  $ET_{VM_k}(T_i)$

$$ET_{VM_k}(T_i) = \frac{Size_{T_i}}{Speed_{VM_k}} .$$

The size of an output data file  $DataFile_{T_i}$  and the typical bandwidth  $\beta$  may be used to compute the communication time  $TT_{E_{ij}}$ :

$$TT_{E_{ij}} = \frac{DataFile_{T_i}}{\beta} .$$

If a task is appear as root task or all parent tasks are on the same VM then communication time is zero.

Step 3.2. Calculate Execution start time  $ST_{T_i}$  and finish time  $FT_{T_p}$  now.  $ST_{T_i}$  is an estimated time to start the execution. It is equal to acquisition delay if the task is appear as root node, otherwise

$$ST_{T_i} = \{Max\{Avail\{VM_k\}, Max_{T_p}\}FT_{T_p} + TT_{E_{pi}}\} .$$

Here  $Avail(VM_k)$  is the time of VMk when it is ready to execute a new task and the VM's performance variation is denoted by PerVar.  $FT_{T_p}$  indicates completion time of parent's task.

$$Avail(VM_k) = \left\{ ST_{T_i} + \left\{ \frac{ET_{VM_k}(T_i)}{(1 - PerVar)} \right\} \right\}.$$

$FT_{T_i}$  is the time indicates to finish the execution.

$$FT_{T_i} = \left\{ ST_{T_i} + \left\{ \frac{ET_{VM_k}(T_i)}{(1 - PerVar)} \right\} \right\}.$$

Step 3.3. Now compute TET and TEC as given below:

$$\text{If } TET \leq D, TET = \{FT(t_i)\},$$

$$TEC = \sum_{n=1}^{VM_n} C_{type(VM_k)} * \left[ \frac{VMno\_ET - VMno\_ST}{TimeInterval} \right].$$

Equation (1) to Equation (8) are from literature [19], [20].

**Step 4.** After computing the fitness of the chromosome, the tournament-based algorithm is used to select the best two individuals for further crossover.

**Step 5.** A two-point crossover is used as depicted in Fig. 3.

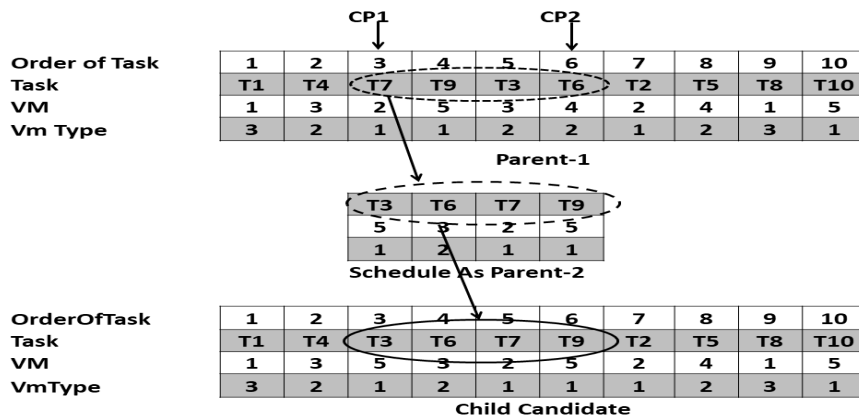


Fig. 3. Crossover Operation

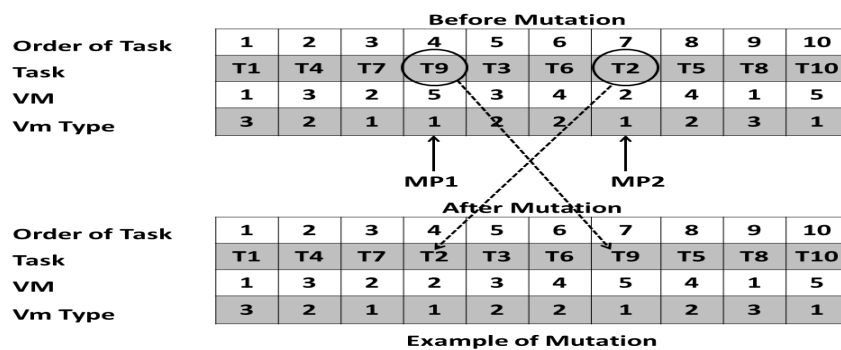


Fig. 4. Mutation Operation

**Step 6.** Apply mutation operations as illustrated in Fig. 4. Now check the dependency constraint on it.

If a new individual follows the dependency constraint, then it is accepted, otherwise it is discarded.

**Step 7.** If the fittest solution meets our objectives under the user-defined constraint, then stop the iteration; otherwise, continue it from step 3 after replacing the least fit candidate with the better new solution.

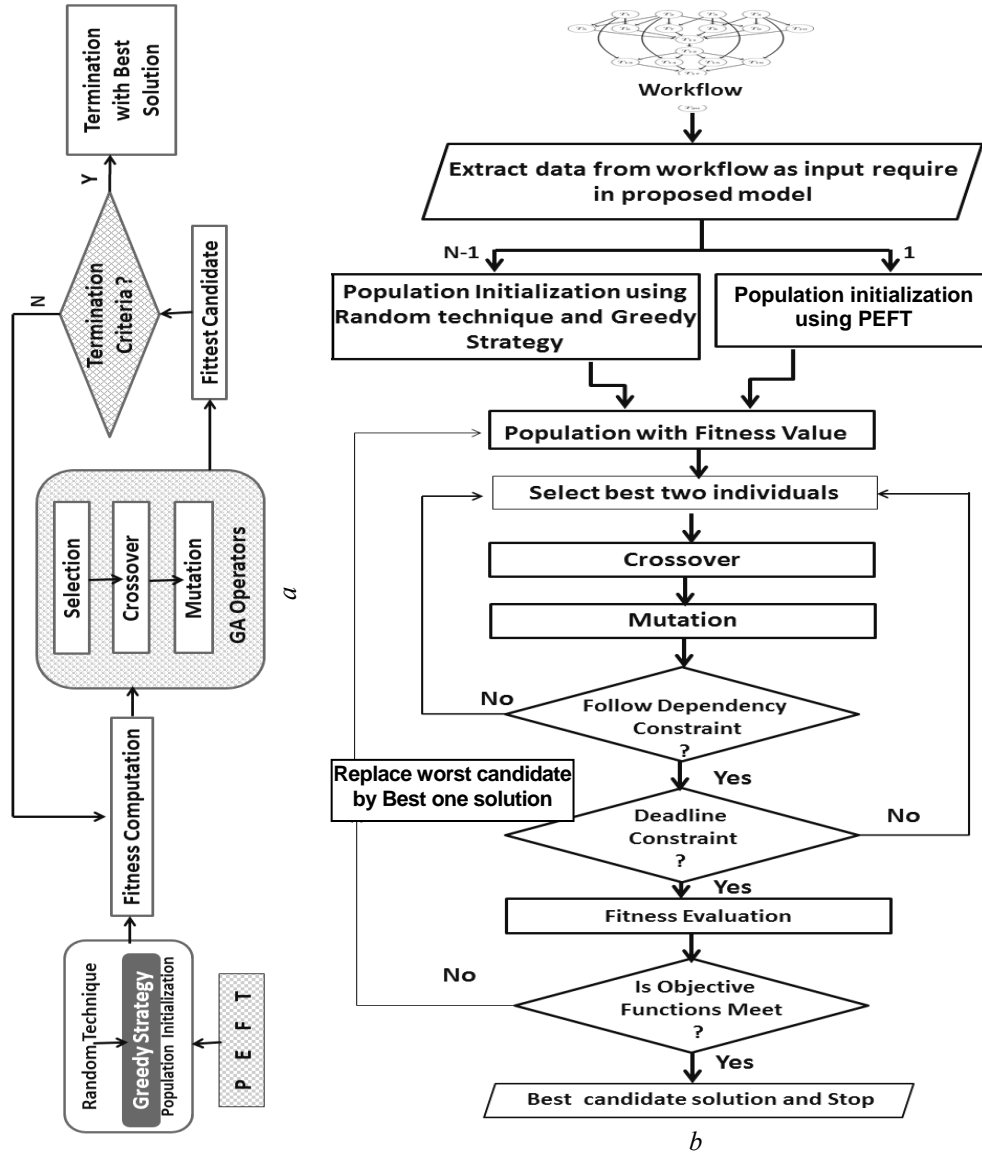


Fig. 5. Block Diagram of Proposed model CHGA (a); flowchart of Proposed model CHGA (b)

A picture is worth a thousand words, that's why we depict our proposed model CHGA through a block diagram and flowchart, as shown in Fig. 5, a and Fig. 5, b respectively.

### A Glance on PEFT

The PEFT [30] consists of two stages: a task prioritization phase that identifies priority of task and a VM selection step that determines the optimum VM for executing the present job. Both stages are centered on OCT. By computing an OCT and retaining quadratic time complexity, this algorithm can forecast. Earliest Fin-

ish Time (EFT) of a node  $n$  on processor  $p$  is sum of earliest start time and computation time of a node  $n$  on processor  $p$ . Illustrated in Fig. 6.

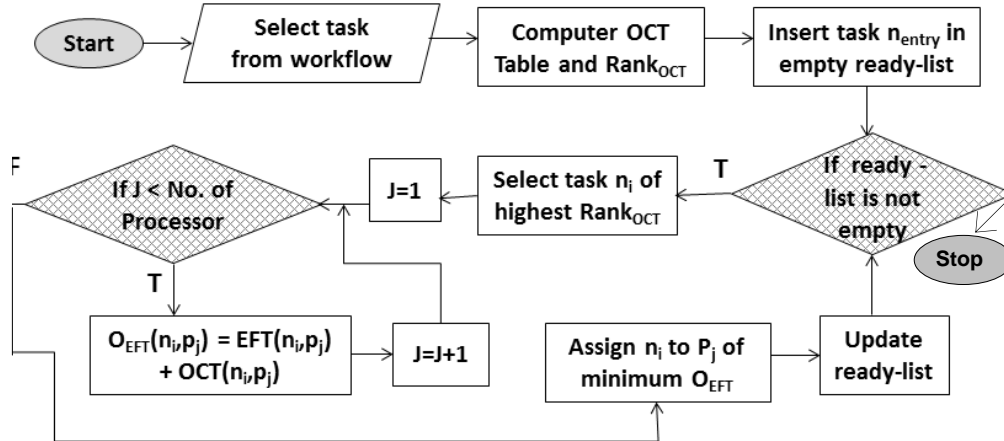


Fig. 6. PEFT Strategy

## PERFORMANCE EVALUATION

### Baseline Algorithm

In the current era, ACO and PCO are buzzwords. Both meta-heuristic algorithms are inspired by the natural process of resolving NP-hard problems like optimization. That’s why we used them as baseline algorithms as they contributed to solving the same problem addressed here. Except these we used CEGA [20] and CLGA [10] as baseline algorithms.

Pheromone-based communication in an ant is to find the best solution. Initially, all the routes have the same probability of selection, i.e., ‘no bias’ due to the same or no pheromone. A local update rule is applied when the ant constructs the route, i.e., solution. Longer pathways vaporise or disintegrate more quickly than shorter ones do. Shorter pathways therefore accumulate more pheromones over time. Pheromone’s quantity is responsible for indirect communication, which is known as ‘stigmergy’ [18]. When all the ants have completed their routes, then a global update operation is performed. Now the selection of the path is biased and the best ant is allowed to update the pheromone by the pseudo-random-proportional rule [18]. We can understand ACO from Fig. 7 and PSO from Fig. 8.

Particle Swarm Optimization was first introduced by Kennedy and Eberhart [22]. In this instance, swarm stands for the population, and particle for a potential solution. Each particle is first assigned a random coordinate. The objective function, or the distance between the particle’s present position and the food, is used to evaluate performance. PBEST indicates the local best position of a particle, whereas refers to the velocity constant. By updating the velocity and position of the particle, a global optimum solution can be achieved. We keep this process going until we get our objective or reach our maximum iterations [22]. This is depicted in Fig. 8. As baseline algorithms we used ACO, PSO, CEGA [20] and CLGA [10].



```

Input: m, α, β, ρ, τ
Output: PBEST
InitPheromone(τ)
PBEST ← NULL
While StopCondition() do
    S_iter ← φ
    For i=1 to m do
        Si ← ConstructSolution(τ)
        If Si is valid output then
            Si ← LocalSearch(Si)
            If (Obj_f(Si) < Obj_f(PBEST)) OR (PBEST = NULL)
                PBEST ← Si
            end
        end
        S_iter ← S_iter ∪ {Si}
    end
    UpdatePheromone(τ, S_iter, PBEST)
end
Return PBEST
    
```

Fig. 7. ACO Algorithm

```

Input: C1, C2, N
Output: GBEST
S ← InitParticle()
PBEST ← NULL
While StopCondition() do
    For each particle Pi in S do
        // Update Local Best
        If f(Pi) < f(PBEST) then
            PBEST ← Pi
        end
    end
    GBEST ← UpdateGlobalBest()
end
For each particle Pi in S do
    UpdateVelocity()
    UpdatePosition()
end
Return GBEST
    
```

Fig. 8. PSO Algorithm

## Experimental Setup

We used four types of scientific workflows: Montage, Cybershake, LIGO, and Epigenomics as benchmarks, where the size of the workflow is 50 nodes, 100 nodes, and 500 nodes approximately.

We have implemented the proposed model CHGA in a JAVA-based robust environment and concluded the result after executing each type of workflow 30 times. The accuracy of the obtained result varies by about ±5. As mentioned in Table 1, we considered 5 types of VMs as specification [31]. We assumed 20 kbps average bandwidth as proposed by Amazon Elastic Block Store (EBS) [32]. A thorough analysis of the literature [33],[34] is beneficial in deciding on various parameters.

There are 3 levels of deadline constraints: hard, crunch, and soft, which are considered in our experiment.

Deadline

$$D = (\beta + 1) \min ET ()W_i .$$

For hard deadline range of  $\alpha$  :  $0 \leq \alpha \leq 1.2$  .

For crunchy deadline range of  $\alpha$  :  $1.2 < \alpha \leq 2.8$  .

For soft deadline range of  $\alpha$  :  $2.8 < \alpha \leq 4.4$  .

Here  $\alpha$  indicates step length, whose value is 0.4.

**Table 1.** Configuration of VM in our practical approach

VM Types	m1.Small	m1.Large	m1.Xlarge	c1.Medium	C1.Xlarge
Processing Capacity (GFLOPS)	4.4	17.6	35.2	22	88
ECUs (Speed)	1	4	8	5	20
Cores	1	2	4	2	8
Memory (GB)	1.7	7.5	15	1.7	7
Disk(GB)	160	850	1690	350	1690
Cost /Hr. (\$)	0.04	0.16	0.32	0.2	0.8

## RESULT AND ANALYSIS

### Evaluation of Deadline Constraint

Our suggested CHGA is evaluated and compared using baseline algorithms in order to fulfill our goal within a user-defined deadline, as depicted in Table 2 and Fig. 9. The hit rate of our proposed CHGA is better than that of other baseline algorithms, which represents its robustness. The capacity of PEFT to predict the impact of scheduling the all children task of the current parent task reduced the makespan of workflow and improved the hit rate of CHGA.

**Table 2.** Analysis of Hit Rate based on deadline

Deadline	Algorithm	Montage	Cybershake	LIGO	Epigenomics
Hard	CHGA	96.3002	94.0645	93.0989	92.3004
	ACO	53.9809	58.2309	52.0051	57.4506
	PSO	69.0989	67.9882	68.0898	69.1216
	CEGA	92.3433	88.4844	88.5034	83.4908
	CLGA	95.5022	91.4788	91.4602	88.0223
Crunch	CHGA	99.8956	99.8002	99.7444	99.8288
	ACO	72.0989	73.0899	71.9004	74.0112
	PSO	79.0767	80.3503	81.0302	78.9704
	CEGA	99.5011	99.6202	99.5002	99.6055
	CLGA	99.5067	99.7601	99.5676	99.7388
Soft	CHGA	99.8876	99.7909	99.7708	99.8092
	ACO	78.0998	76.0038	77.0902	78.2312
	PSO	83.4534	82.0034	85.7801	86.5709
	CEGA	99.6767	99.7022	99.6081	99.5003
	CLGA	99.6878	99.7803	99.7003	99.7099

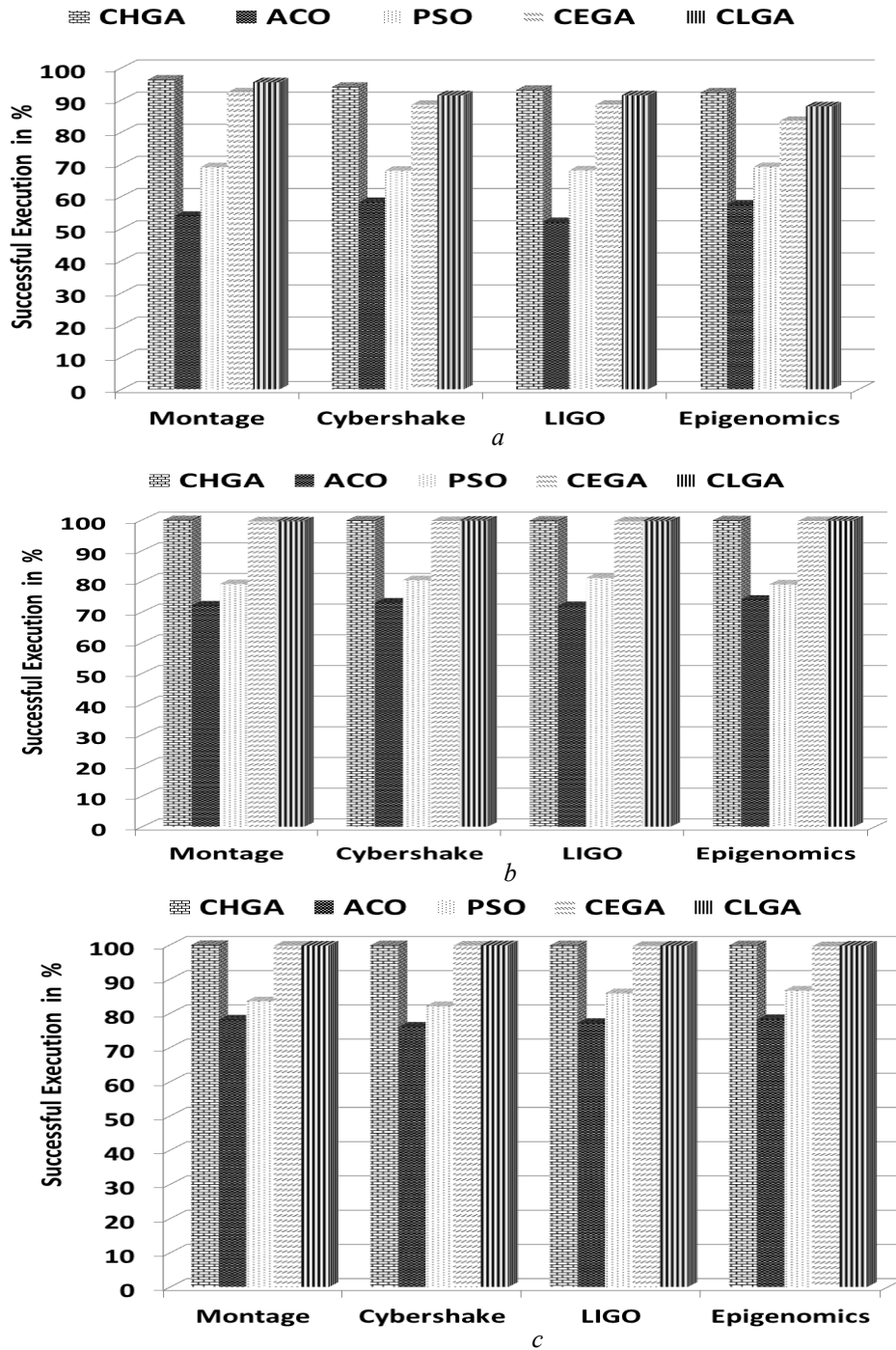


Fig. 9. Analysis under: Hard Deadline Constraint (a); Crunch Deadline Constraint (b); Soft Deadline Constraint (c)

### Load-Balance Evaluation

Greedy strategy plays an important role in load balance. Finding a virtual machine with a low load is important before we allocate a task  $T_i$  to an individual. In order

to manage the load balance, we map the task  $T_i$  with  $VM_i$  using the greedy technique after identifying  $VM_i$  with minimum load.

The capacity of  $VMC_i$  can be calculated as given by Equation, where the number of processing elements is  $PE_{num}$ :

$$VMC_i = PE_{num} \times PE_{mips} .$$

All  $VMC_i$  are collectively known as Virtual Machine Capacity (VMC), and  $m$  is the total number of  $VM_s$  :

$$VMC = \sum_{i=1}^m VMC_i .$$

Load  $L_i$  on a  $VM_i$  is as Equation.

Total load TL is as Equation

$$TL = \sum_{i=1}^m L_i .$$

Load capacity per unit is  $LC_{pu}$  as Equation

$$LC_{pu} = \frac{TL}{VMC} .$$

Threshold value  $TH_i$  is as Equation

$$TH_i = LC_{pu} \times VMC_i .$$

The threshold value  $TH_i$  is compared with the load of  $VM_i$  to determine the status of  $VM_i$ , i.e., under-loaded, balanced or over-loaded. The result of our experiment shows that with ACO, VM1 is overloaded by +82% and VM3 is under-loaded by -58%. In contrast, with PSO, VM5 is overloaded by + 69% and VM4 is under-loaded by -63%, as shown in Fig. 10. Our model CHGA exhibited better load-balance compare to ACO, PSO, and CEGA, which denotes the robustness of CHGA. When we used the proposed CHGA, VM4 was overloaded by +26%, while VM3 was under-loaded by -12%. Illustrated in Fig. 10.

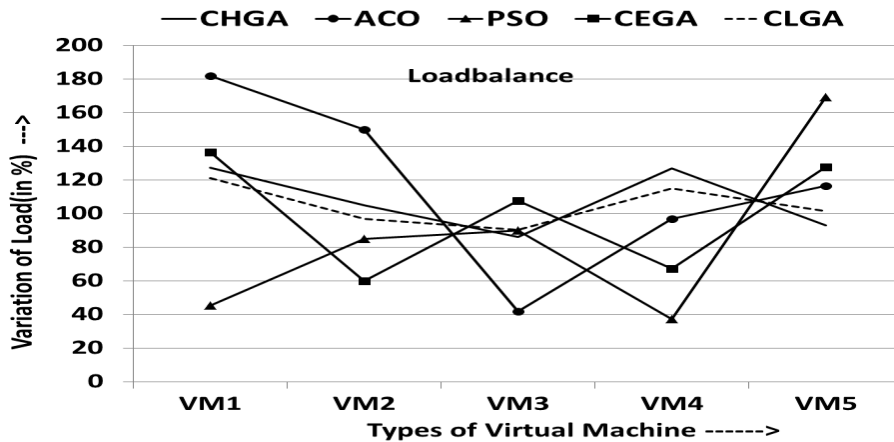


Fig. 10. Comparatively analysis of Load Balance

### Cost and Makespan Evaluation

Our holistic comparison between the baseline and our proposed CHGA is depicted in Fig. 11 – 14. The obtained result of our experiment indicates the robustness of our proposed model CHGA. CHGA is 14.58188%, 11.40224%, 11.75306%, and 9.78841% cheaper than standard ACO, PSO, CEGA, and CLGA, respectively. CHGA's average makespan is 34.73619%, 31.48127%, 5.71553%, and 9.73710% lower than standard ACO, PSO, CEGA, and CLGA, respectively.

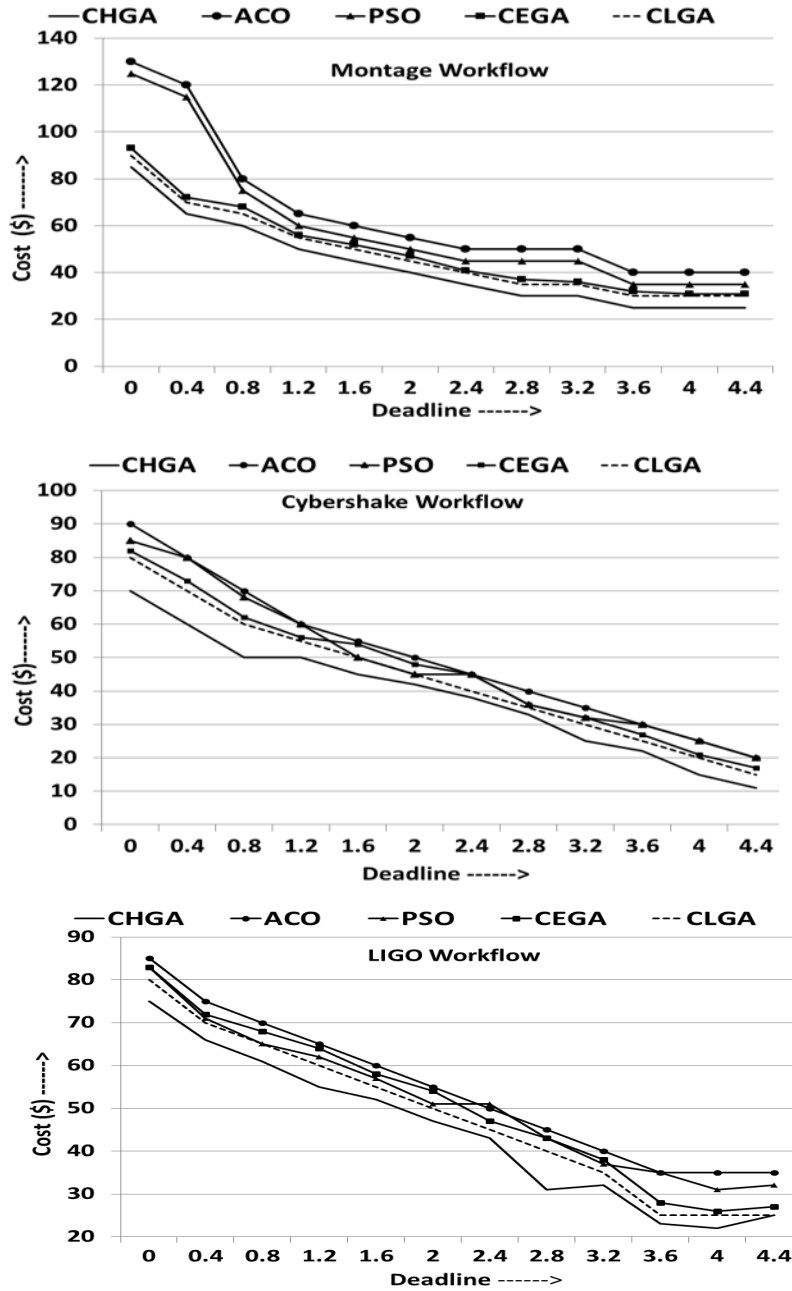


Fig. 11. Comparatively analysis of Cost. *Began*

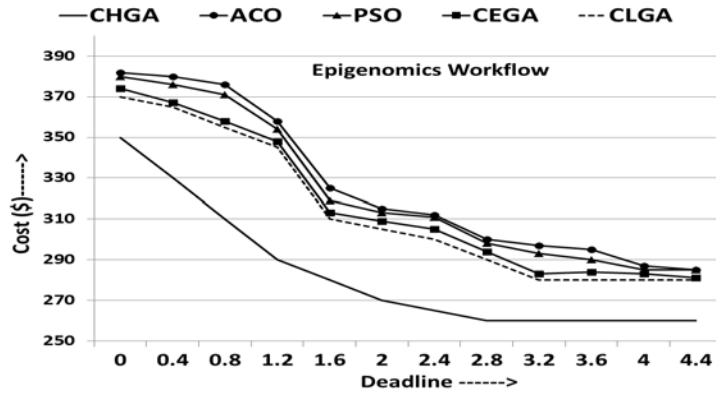


Fig. 12. Comparatively analysis of Cost. *Continued*

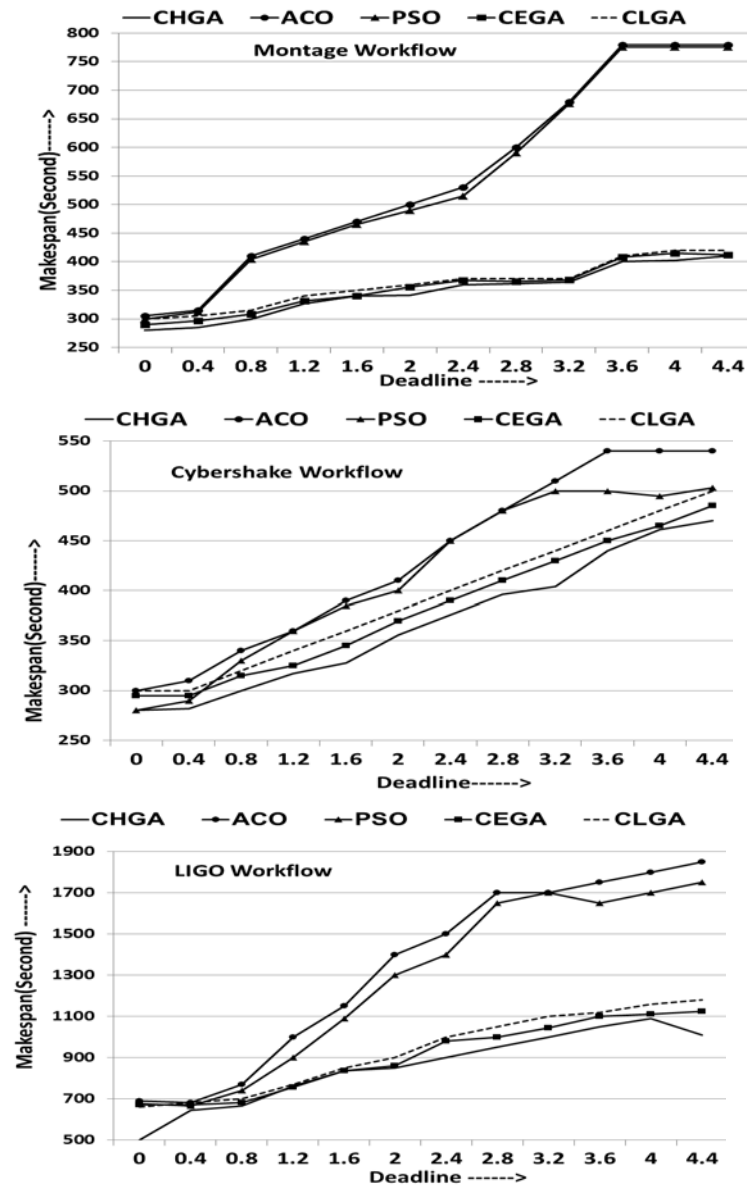


Fig. 13. Comparatively analysis of Makespan. *Began*

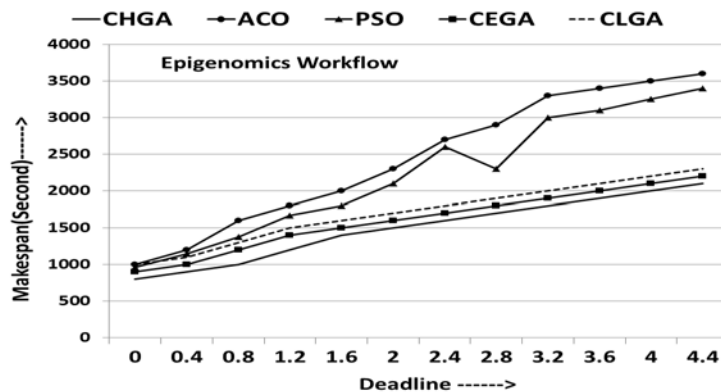


Fig. 14. Comparatively analysis of Makespan. *Continued*

The capacity of PEFT to predict the impact of scheduling the all children task of the current parent task reduced the makespan of workflow and improved the execution cost in term of minimization in our proposed model. The obtained result of our experiment indicates the robustness of our proposed model CHGA. CHGA is 14.58188%, 11.40224%, 11.75306%, and 9.78841% cheaper than standard ACO, PSO, CEGA, and CLGA, respectively. CHGA's average makespan is 34.73619%, 31.48127%, 5.71553%, and 9.73710% lower than standard ACO, PSO, CEGA, and CLGA, respectively.

## DISCUSSION

Because the best schedules take into account both the gain in a sequence of tasks as well as the immediate gain in processing time, we observed that the best meta-heuristic schedules could not be achieved if we adhered to the conventional strategy of selecting processors based only on current task execution time, so we used the PEFT strategy during population initialization, which helps to reduce the makespan. Its capacity to predict the impact of scheduling all child tasks of the current parent task This attribute allows one to make the perfect decision when selecting the perfect virtual machine. We also took into account the time it takes for a VM to boot up and performance fluctuations, both of which have an influence on computation time and execution cost. These statements are verified by the obtained results of our experiments, which indicate the robustness of our proposed model CHGA. CHGA is 14.58188%, 11.40224%, 11.75306%, and 9.78841% cheaper than standard ACO, PSO, CEGA, and CLGA, respectively. CHGA's average makespan is 34.73619%, 31.48127%, 5.71553%, and 9.73710% lower than standard ACO, PSO, CEGA, and CLGA, respectively.

We also applied the greedy strategy during the initialization of the population, which plays an important role in load balancing among VMs. When we used the proposed CHGA, VM4 was overloaded by +26%, while VM3 was underloaded by -12%, which shows our model CHGA is better in load-balancing compared to ACO, PSO, and CEGA.

## CONCLUSIONS AND FUTURE WORK

To schedule scientific workflow, we introduced our meta-heuristic, cost-effective, load-balanced hybrid evolutionary method. To balance the load among VMs in a

heterogeneous environment, an effective encoding approach with a greedy strategy is used. We also employed the PEFT technique to make our algorithm more cost-effective. Under a user-defined deadline, we considered three parameters: makespan, computation cost, and load balance, and rigorously tested four types of scientific workflows with varied task sizes. Our experimental results proved that the proposed CHGA algorithm's performance is better than the ACO, PSO, CEGA, and CLGA in minimizing the computing cost and execution time as well as balancing the load among virtual machines. CHGA is 17.48570%, 15.30489%, 11.75306%, and 9.78841% cheaper than standard ACO, PSO, CEGA, and CLGA, respectively. CHGA's average makespan is 34.73619%, 31.48127%, 5.71553%, and 9.73710% lower than standard ACO, PSO, CEGA, and CLGA, respectively. In the future, we will consider the dynamic nature of workflow with the latest meta-heuristic algorithms like Cuckoo search, Firefly, Lion, and Jaya, etc.

### CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

### REFERENCES

1. S. Ibrahim, B. He, and H. Jin, "Towards pay-as-you-consume cloud computing," *Proc. 2011 IEEE Int. Conf. Serv. Comput. SCC 2011*, pp. 370–377. doi: 10.1109/SCC.2011.38.
2. R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Futur. Gener. Comput. Syst.*, 25(6), pp. 599–616, 2009. doi: 10.1016/j.future.2008.12.001.
3. A.S. Kulik, A.G. Chukhray, and O.V. Havrylenko, "Information technology for creating intelligent computer programs for training in algorithmic tasks. Part 1: mathematical foundations," *System Research & Information Technologies*, no. 4, 2021. doi: 10.20535/SRIT.2308-8893.2021.4.02
4. S. Singhal and J. Grover, "Hybrid biogeography algorithm for reducing power consumption in cloud computing," *2017 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2017*, pp. 121–124. doi: 10.1109/ICACCI.2017.8125827.
5. S. Singhal and J. Patel, "Load balancing scheduling algorithm for concurrent workflow," *Comput. Informatics*, 37(2), pp. 311–326, 2018. doi: 10.4149/cai\_2018\_2\_311.
6. G. Dalin and V. Radhamani, "IRIAL-an improved approach for VM migrations in cloud computing," *International Journal of Advanced Technology and Engineering Exploration*, 2018 July 1, 5(44), pp. 165–171.
7. J. Yu and R. Buyya, "A taxonomy of workflow management systems for Grid computing," *J. Grid Comput.*, 3(3–4), pp. 171–200, 2005. doi: 10.1007/s10723-005-9010-8.
8. D. Malhotra, "An adaptive threshold policy for host overload detection in cloud data centre," *International Journal of Advanced Technology and Engineering Exploration*, 2021 Oct 1, 8(83), pp. 1315–1335.
9. S.K. Bothra and S. Singhal, "Nature-inspired metaheuristic scheduling algorithms in cloud: A systematic review," *Sci. Tech. J. Inf. Technol. Mech. Opt.*, 21(4), pp. 463–472, 2021. doi: 10.17586/2226-1494-2021-21-4-463-472.
10. S.K. Bothra, S. Singhal, and H. Goyal, "Deadline-constrained cost-effective load-balanced improved genetic algorithm for workflow scheduling," *Int. J. Inf. Technol. Web Eng.*, 16(4), pp. 1–34, 2021. doi: 10.4018/IJITWE.2021100101.



11. L.F. Bittencourt, R. Sakellariou, and E.R.M. Madeira, "DAG scheduling using a look-ahead variant of the heterogeneous earliest finish time algorithm," *Proc. 18th Euro-micro Conf. Parallel, Distrib. Network-Based Process, PDP 2010*, pp. 27–34. doi: 10.1109/PDP.2010.56.
12. V.M. Sineglazov, K.D. Riazanovskiy, and O.I. Chumachenko, "Multicriteria conditional optimization based on genetic algorithms," *System Research & Information Technologies*, no. 3, pp. 89–104, 2020. doi: 10.20535/SRIT.2308-8893.2020.3.07
13. Dorigo and C. Blum, "Ant colony optimization theory: A survey," *Theoretical Computer Science*, 344, pp. 243–278, 2005. doi: 10.1016/j.tcs.2005.05.020.
14. S.K. Patel and A.K. Sharma, "Improved PSO based job scheduling algorithm for resource management in grid computing," *International Journal of Advanced Technology and Engineering Exploration*, 2019 May 1, 6(54), pp. 152–161.
15. D. Karaboga and B. Akay, "A modified Artificial Bee Colony (ABC) algorithm for constrained optimization problems," *Appl. Soft Comput. J.*, 11(3), pp. 3021–3031, 2011. doi: 10.1016/j.asoc.2010.12.001.
16. Z. Wu, Z. Ni, and L. Gu, "A Revised Discrete Particle Swarm Optimization for Cloud Workflow Scheduling," *Proceedings of the 2010 International Conference on Computational Intelligence and Security*, pp. 184–188. doi: 10.1109/CIS.2010.46.
17. S. Pandey, L. Wu, S.M. Guru, and R. Buyya, "A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments," *Proc. Int. Conf. Adv. Inf. Netw. Appl. AINA, 2010*, pp. 400–407. doi: 10.1109/AINA.2010.31.
18. F. Engineering, "Scheduling Workflow in Cloud Computing Based on Ant Colony Optimization Algorithm," *2013 Sixth International Conference on Business Intelligence and Financial Engineering (BIFE)*, pp. 57–61. doi: 10.1109/BIFE.2013.14.
19. Z. Chen and K. Du, "Deadline Constrained Cloud Computing Resources Scheduling for Cost Optimization based on Dynamic Objective Genetic Algorithm," *Evolutionary Computation (CEC), 2015 IEEE Congress on, IEEE, 2015*, pp. 708–714.
20. J. Meena, M. Kumar, and M. Vardhan, "Cost Effective Genetic Algorithm for Workflow Scheduling in Cloud Under Deadline Constraint," *IEEE Access*, 4, pp. 5065–5082, 2016. doi: 10.1109/ACCESS.2016.2593903.
21. M. Mollajafari and H.S. Shahhoseini, "A cost-optimized GA-based heuristic for scheduling time-constrained workflow applications in infrastructure clouds using an innovative feasibility-assured decoding mechanism," *J. Inf. Sci. Eng.*, 32(6), pp. 1541–1560, 2016. doi: 10.1688/JISE.2016.32.6.8.
22. Gupta, V. Gajera, P.K. Jana, and I.S. Member, "An Effective Multi-Objective Workflow Scheduling in Cloud Computing: A PSO based Approach," *Ninth International Conference on Contemporary Computing*, 2016.
23. Y. Moon, H. Yu, J.M. Gil, and J. Lim, "A slave ants based ant colony optimization algorithm for task scheduling in cloud computing environments," *Human-centric Comput. Inf. Sci.*, 2017. doi: 10.1186/s13673-017-0109-2.
24. A. You, M.A.Y. Be, and I. In, "Task scheduling based on ant colony optimization in cloud environment," *AIP Conference Proceedings*, vol. 1834, iss. 1, 040039, 2017. doi: 10.1063/1.4981635.
25. B. Xiang, B. Zhang, and L. Zhang, "Greedy Ant: Ant Colony System-Inspired Workflow Scheduling for Heterogeneous Computing," *IEEE Access*, 2017. doi: 10.1109/ACCESS.2017.2715279.
26. S.K. Sonkar and M.U. Kharat, "Load prediction analysis based on virtual machine execution time using optimal sequencing algorithm in cloud federated environment," *Int. J. Inf. Technol.*, 11(2), pp. 265–275, 2019. doi: 10.1007/s41870-019-00282-1.
27. C. Yan, M.X. Li, and W. Liu, "Application of improved genetic algorithm in function optimization," *J. Inf. Sci. Eng.*, 35(6), pp. 1299–1309, 2019. doi: 10.6688/JISE.201911\_35(6).0008.

28. Al-Azzawi, "Evaluation of Genetic Algorithm Optimization in Machine Learning," *J. Inf. Sci. Eng.*, 36(2), 2020.
29. A. Belgacem and K. Beghdad-Bey, "Multi-objective workflow scheduling in cloud computing: trade-off between makespan and cost," *Cluster Computing*, 25(1), pp. 579–595, 2022.
30. H. Arabnejad and J.G. Barbosa, "List Scheduling Algorithm for Heterogeneous Systems by an Optimistic Cost Table," *IEEE Transactions on Parallel and Distributed Systems*, 25(3), pp. 682–694, 2014.
31. Z. Hill and M. Humphrey, "A quantitative analysis of high performance computing with Amazon's EC2 infrastructure: The death of the local cluster?" *Proc. IEEE/ACM Int. Work. Grid Comput.*, pp. 26–33, 2009. doi: 10.1109/GRID.2009.5353067.
32. *Amazon Elastic Block Store*. Available: <https://aws.amazon.com/ebs/>. Accessed 22 July, 2020.
33. M. Naderan, "Review methods for breast cancer detection using artificial intelligence and deep learning methods," *System Research & Information Technologies*, no. 1, 2021. doi: 10.20535/SRIT.2308-8893.2021.1.08
34. I.V. Beyko, J.V. Spivak, and O.V. Furtel, "Generalized solutions of optimal control problems," *System Research & Information Technologies*, no. 4, 2020, pp. 104–114. doi: 10.20535/SRIT.2308-8893.2020.4.08.

Received 15.08.2022

#### INFORMATION ON THE ARTICLE

**Sandeep Kumar Bothra**, ORCID: 0000-0003-0555-569X, Manipal University Jaipur, (Raj.), India, e-mail: bothrajain@gmail.com

**Sunita Singhal**, ORCID: 0000-0003-2462-8102, Manipal University Jaipur, (Raj.), India, e-mail: sunita.singhal@jaipur.manipal.edu

**Hemlata Goyal**, ORCID: 0000-0003-1344-0921, Manipal University Jaipur, (Raj.), India, e-mail: hemlata.goyal@jaipur.manipal.edu

#### ЕКОНОМІЧНО ЕФЕКТИВНИЙ ГІБРИДНИЙ ГЕНЕТИЧНИЙ АЛГОРИТМ ПЛАНУВАННЯ РОБОЧОГО ПРОЦЕСУ В ХМАРІ / Сандіп Кумар Бовра, Суніта Сінгхал, Хемлата Гоял

**Анотація.** Хмарні обчислення відіграють значну роль у способі життя кожного, щільно пов'язуючи спільноти, інформацію та торги по всьому світу. Розпізнавання оптимального рішення для планування робочих процесів у хмарі є складною сферою через його NP-жорсткий характер. Запропоновано гібридний метаевристичний економічно ефективний збалансований за навантаженням підхід до планування робочого процесу в гетерогенному середовищі. Модель ґрунтується на генетичному алгоритмі, інтегрованому з прогнозом найбільш раннього часу фінішу (PEFT), щоб мінімізувати makepan. Замість призначення завдання випадковим чином на віртуальній машині застосовуємо жадібну стратегію, яка відводить завдання на віртуальну машину з найменш навантаженим. Після завершення операції мутації перевіряємо обмеження залежності замість кожної операції кросовера, що дає кращий результат. Запропонована модель включає в себе дисперсію продуктивності віртуальної машини, а також затримку придбання, яка поступається мінімальній вартості makepan і computing. Одним з найбільш приголомшливих аспектів економічно ефективного гібридного генетичного алгоритму (CHGA) є його здатність передбачати, створюючи оптимістичну таблицю витрат (ОСТ), зберігаючи квадратичну складність часу. На основі результатів ретельних експериментів над деякими показниками робочого процесу в реальному світі та всебічного аналізу деяких нещодавно успішних алгоритмів планування отримано висновок, що продуктивність запропонованої CHGA є мелодійною.

**Ключові слова:** хмарні обчислення, економічно вигідні, генетичний алгоритм, метаевристичний алгоритм, прогнозування раннього часу оброблення, планування робочого процесу.

## MULTI-STEP PREDICTION IN LINEARIZED LATENT STATE SPACES FOR REPRESENTATION LEARNING

A. TYTARENKO

**Abstract.** In this paper, we derive a novel method as a generalization over LCEs such as E2C. The method develops the idea of learning a locally linear state space by adding a multi-step prediction, thus allowing for more explicit control over the curvature. We show that the method outperforms E2C without drastic model changes which come with other works, such as PCC and P3C. We discuss the relation between E2C and the presented method and derive update equations. We provide empirical evidence, which suggests that by considering the multi-step prediction, our method – ms-E2C – allows learning much better latent state spaces in terms of curvature and next state predictability. Finally, we also discuss certain stability challenges we encounter with multi-step predictions and how to mitigate them.

**Keywords:** representation learning, learning controllable embedding, reinforcement learning, latent state space.

### INTRODUCTION

One of the most challenging problems which the field reinforcement learning faces is learning autonomous agents capable of control in Markov Decision Processes (MDP) with complex state and action spaces. For instance, complications may arise from large action spaces [1], limited ability to interact with an environment [2], partial observability (POMDP) [3, 4], etc. Optimizing a decent policy takes a lot of samples, usually requires online interactive learning and neural networks capable of processing higher dimensional observations with large number of trainable parameters [5, 6].

There are various algorithms which try to deal with the problem of sample inefficiency, or limited amount of data. Model-based reinforcement learning algorithms [7–9] try to achieve sample efficiency by approximating transition dynamics of an MDP in online or offline mode. Offline reinforcement learning methods [2, 10] strive to extract as much useful information from limited offline data as possible, in order to learn a policy applicable to online regimes as well.

Another algorithmic framework – Learning Controllable Embedding (LCE) – approaches this problem by learning a lower dimensional latent state space and using simpler control algorithms, like iLQR [11], to perform control in this latent space. The challenge here is to make sure that the learned latent space has simpler structure (i.e. next states are easier to predict).

Some particular instances of this framework are described in [9, 12–14]. The idea of E2C [12] is to learn a locally-linear latent space, so that algorithms like LQG could be used for goal-reaching tasks. PCC [13] tries to fix some of the issues encountered in E2C by deriving losses which allow for explicit minimization of latent space's curvature. P3C [14] improves upon PCC mainly by replacing reconstruction loss, needed to make sure the learned state space carries enough

information to generate (i.e. decode) observations from latent states. P3C uses predictive coding instead.

In this paper, we seek an alternative approach to enforce lower latent space's curvature and predictability. We generalize E2C by considering multiple transitions at a time, making sure the local linearity is not just preserved between neighbouring states. We inherit the idea of minimization of a joint log likelihood of a transition, generalize it to multiple transitions, and derive a variational bound for further minimization. We then compare the results with LCE approaches and demonstrate a visual representation of learned latent state spaces for a benchmark common among LCE papers.

## PRELIMINARIES

We denote a **Markov Decision Process** (MDP)  $M$  as a tuple  $(S, A, r, T)$ , where  $S$  — state space;  $A$  — action space;  $r: S \times A \rightarrow \mathbb{R}$  — reward function;  $T = P(s_{t+1} | s_t, a_t)$  — probability of state  $s_{t+1}$  given current state  $s_t$  and action taken  $a_t$ .

A state of an MDP is a sufficient statistics for a transition kernel, possessing a Markov property.

A task of **Reinforcement Learning** algorithm is for a given MDP  $M$  find a policy  $\pi$ , such that it maximizes the expected return. We are interested in a discounted return objective:

$$\pi^* = \arg \max_{\pi} E_{\tau \sim p(\cdot)} \sum_{t=0}^{\infty} \gamma^t r(s_t, a_t),$$

where  $\tau$  denotes a trajectory  $(x_0, a_0, x_1, a_1, \dots)$  obtained by sampling actions using a stochastic policy  $\pi$ .

In particular, we consider a specific class of Reinforcement Learning algorithms – **Model Based Reinforcement Learning** [7, 8]. Algorithms of this kind usually possess higher sample efficiency, but they involve some sort of an approximation of a transition kernel. LCE algorithms involve parametric models (i.e. neural nets) to learn a good model with desirable properties, like linearity and predictability. This allows to use even the simplest Model-based control (and RL) algorithms like iLQR [11].

## THE MULTI-STEP EMBED TO CONTROL MODEL

Consider an internal transition dynamics of an MDP  $M = (S, A, r, T)$ :

$$s_{t+1} = f(s_t, a_t) + \omega, \quad \omega \sim P_{\omega}.$$

As we discussed previously, a function  $f$  may be highly nonlinear, thus being tricky to optimize with model-based RL or control algorithms. LCE approaches therefore try to learn a mapping from a state space  $S$  to some latent space  $Z$  such that its latent dynamics

$$z_{t+1} = \hat{f}(z_t, a_t) + \hat{\omega}$$

has some desired properties like local linearity, low curvature, predictability, etc.

In order to learn the mapping  $Q_\phi : S \rightarrow Z$ , Variational Inference framework is employed to derive a tractable algorithm of maximization a likelihood of known data points under the mapping we want to learn.

### Optimization problem

As follows from the Fig. 1, we consider a dataset

$$D = \{(s_t, a_t, s_{t+1}, a_{t+1}, s_{t+2}, \dots, a_{t+K-1}, s_{t+K}) \mid i = 1, \dots, N\}$$

containing samples from real trajectories gathered before training. To follow and generalize [12], we define

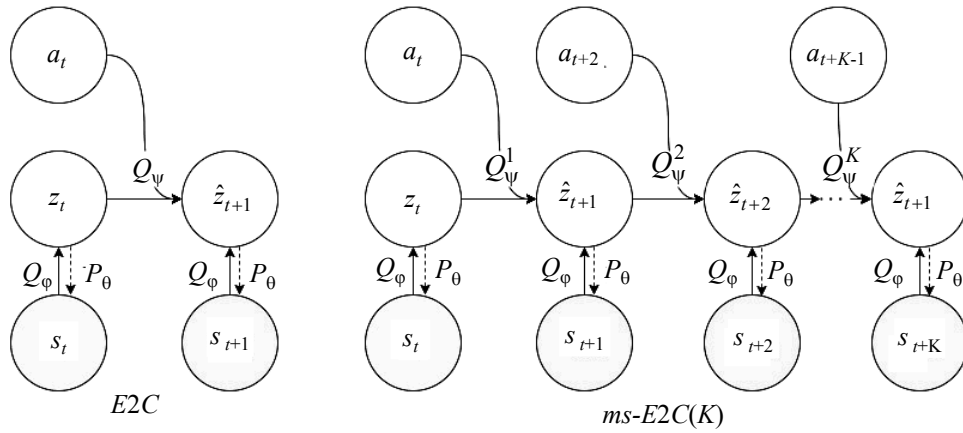


Fig. 1. Graphical models for E2C and ms-E2C(K): dashed lines – state reconstruction process

$$Q_\phi = P_\phi(z_t | s_t)$$

as a generative model which for a given state  $s_t$  specifies a distribution over the latent space  $Z$ . Basically, it plays a role of the mapping from  $S$  to  $Z$  parametrized with a parameter vector  $\phi$ . And

$$Q_\psi = P_\psi(\hat{z}_{t+1} | \hat{z}_t, a_t)$$

as a generative model which for a given latent state  $z_t$  and an action  $a_t$  predicts the distribution for the next latest state  $z_{t+1}$ . The model is also parametrized with a parameter vector  $\psi$ . Also, we denote

$$Q_\psi^j = P_\psi(\hat{z}_{t+j} | \hat{z}_{t+j-1}, a_t).$$

In order to find  $\phi$  and  $\psi$ , we maximize the likelihood of a dataset of trajectory samples of length  $K$  with respect to the aforementioned parameter vectors:

$$\phi^*, \psi^* = \arg \max_{\phi, \psi} \prod_{i=1}^N P(s_t^i, a_t^i, s_{t+1}^i, a_{t+1}^i, s_{t+2}^i, \dots, a_{t+K-1}^i, s_{t+K}^i).$$

For the sake of readability, we denote  $s_t, \dots, s_{t+K}$  as  $s_{t:t+K}$  and  $a_t, \dots, a_{t+K}$  as  $a_{t:t+K}$ . Thus our objective is:

$$\phi^*, \psi^* = \arg \max_{\phi, \psi} \prod_{i=1}^N P(s_{t:t+K}^i, a_{t:t+K-1}^i).$$

A corresponding graphical model is depicted on Fig. 1.

### Optimization objective

The objective we defined in a previous subsection is known to be intractable and difficult to optimize. Therefore, LCE approaches employ Variational Inference to find a lower bound to the log-likelihood objective. In this section, we derive this bound for the proposed probabilistic model. Variational lower bound:

$$\begin{aligned} & -\log P(s_{t:t+K}, a_{t:t+K-1}) \leq \\ & \leq E_{\substack{z_t \sim Q_\phi \\ \hat{z}_{t+j} \sim Q_\psi^j \\ j=1, \dots, K}} \left[ -\sum_{j=1, \dots, K} \log P(s_{t+j} | \hat{z}_{t+j}) - \log P(s_t | z_t) \right] + D_{KL}[Q_\phi \| P(Z)] \end{aligned}$$

Here  $D_{KL}[P \| Q]$  denotes Kullback–Leibler divergence functional:

$$D_{KL}[P \| Q] = E_{x \sim P} \log \frac{P(x)}{Q(x)}.$$

### Proof.

$$\begin{aligned} -\log P(s_{t:t+K}, a_{t:t+K-1}) &= -\log \int_{z_t, \hat{z}_{t+1:t+K}} P(s_{t:t+K}, a_{t:t+K-1}, z_t, \hat{z}_{t+1:t+K}) dz_t d\hat{z}_{t+1:t+K} = \\ &= -\log \int_{z_t, \hat{z}_{t+1:t+K}} P(s_{t:t+K} | a_{t:t+K-1}, z_t, \hat{z}_{t+1:t+K}) P(z_t, \hat{z}_{t+1:t+K} | a_{t:t+K-1}) \times \\ & \quad \times P(a_{t:t+K-1}) dz_t d\hat{z}_{t+1:t+K} \\ &= -\log \int_{z_t, \hat{z}_{t+1:t+K}} P(s_{t:t+K} | a_{t:t+K-1}, z_t, \hat{z}_{t+1:t+K}) P(z_t, \hat{z}_{t+1:t+K} | a_{t:t+K-1}) \times \\ & \quad \times P(a_{t:t+K-1}) \frac{Q_\phi}{Q_\psi} dz_t d\hat{z}_{t+1:t+K} \\ &= -\log \int_{z_t, \hat{z}_{t+1:t+K}} P(s_{t:t+K} | a_{t:t+K-1}, z_t, \hat{z}_{t+1:t+K}) \prod_{j=1}^K (Q_\psi^j) P(z_t) \times \\ & \quad \times P(a_{t:t+K-1}) \frac{Q_\phi}{Q_\psi} dz_t d\hat{z}_{t+1:t+K} \\ &= -\log \int_{z_t, \hat{z}_{t+1:t+K}} P(s_t | z_t) \prod_{j=1}^K (Q_\psi^j P(s_{t+j} | \hat{z}_{t+j})) P(z_t) P(a_{t:t+K-1}) \frac{Q_\phi}{Q_\psi} dz_t d\hat{z}_{t+1:t+K} = \\ &= -\log E_{\substack{z_t \sim Q_\phi \\ \hat{z}_{t+j} \sim Q_\psi^j \\ j=1, \dots, K}} \left[ P(s_t | z_t) \prod_{j=1}^K (P(s_{t+j} | \hat{z}_{t+j})) P(a_{t:t+K-1}) \frac{P(z_t)}{Q_\phi} \right] \leq \end{aligned}$$

$$\begin{aligned}
 &\leq E_{\substack{z_t \sim Q_\phi \\ \hat{z}_{t+j} \sim Q_\psi^j \\ j=1, \dots, K}} \left[ - \sum_{j=1, \dots, K} \log P(s_{t+j} | \hat{z}_{t+j}) - \log P(s_t | z_t) - \log \frac{P(z_t)}{Q_\phi} \right] = \\
 &= E_{\substack{z_t \sim Q_\phi \\ \hat{z}_{t+j} \sim Q_\psi^j \\ j=1, \dots, K}} \left[ - \sum_{j=1, \dots, K} \log P(s_{t+j} | \hat{z}_{t+j}) - \log P(s_t | z_t) \right] + D_{KL}[Q_\phi \| P(Z)].
 \end{aligned}$$

### Multi-step embed-to-control model (ms-E2C)

In this section we instantiate a model for learning a latent locally-linear state space. We use a previously derived upper bound for negative loglikelihood over the multi-step trajectory samples. A graphical models for both E2C and ms-E2C( $K$ ) are shown on the Fig. 1.

First, we instantiate **parametric models for encoding and dynamics function** as:

$$Q_\phi = P(z_t | s_t) = N(\mu_\phi(s_t), \Sigma_\phi(s_t)) \text{ — encoder;}$$

$$\mu_\phi(s_t), \Sigma_\phi(s_t) = \text{NeuralNet}(s_t; \phi);$$

$$A_t, B_t, o_t = \text{NeuralNet}(z_t; \psi) \text{ — latent dynamics;}$$

$$Q_\psi^j = P(\hat{z}_{t+j} | \hat{z}_{t+j-1}, a_{t+j-1}) = N(\mu^j_\psi, \Sigma^j_\psi) \text{ — dynamics;}$$

$$\mu^j_\psi = A_t \mu^{j-1}_\psi + B_t a_{t+j-1} + o_t \text{ for } j = 2, \dots, K;$$

$$\mu^j_\psi = A_t \mu_\psi + B_t a_t + o_t \text{ for } j = 1;$$

$$\Sigma^j_\psi = A_t \Sigma^{j-1}_\psi A_t^T + \Sigma_W \text{ for } j = 2, \dots, K;$$

$$\Sigma^j_\psi = A_t \Sigma_\psi A_t^T + \Sigma_W \text{ for } j = 1.$$

Thus, given an optimal model would imply a locally linear latent space, in which curvature (i.e. linearity) is explicitly controlled by changing the number of steps per sample. Choosing a large  $K$  would recover a globally linear model and setting  $K=1$  recovers an E2C model.

As it follows from the figure, ms-E2C is a generalization of E2C, which one recovers by setting  $K=1$ .

We also have to specify a **parametrized decoding model**, which is needed to compute the upper bound, and to enforce a "reconstruction" constraint, introduced in [12] and generalized for our multi-step model:

$$P_\theta = P_\theta(s_t | z_t) = \text{Bernoulli}(p(z_t));$$

$$p(z_t) = \text{NeuralNet}(z_t; \theta);$$

$$P_\theta^j = P_\theta(s_t | \hat{z}_{t+j}) = \text{Bernoulli}(p(\hat{z}_{t+j}));$$

$$p(\hat{z}_{t+j}) = \text{NeuralNet}(\hat{z}_{t+j}; \theta).$$

**Bernoulli distribution** is chosen for comparability with E2C model on shared benchmark MDPs, i.e. Planar, where the original state space consists of black-and-white images of a grid world with white obstacles and a white circle denoting the position of the agent.

**Loss function.** In order to complete the model's specification, we have to provide a loss function optimizable via stochastic gradient descent. In msE2C it consists of three terms: an estimation of the derived upper bound, consistency term, and stability term.

$$L_{upper}(D_i; \phi, \psi, \theta) = E_{\substack{z_t \sim Q_\phi \\ \hat{z}_{t+j} \sim Q_\psi^j \\ j=1, \dots, K}} \left[ - \sum_{j=1, \dots, K} \log P_\theta(s_{t+j} | \hat{z}_{t+j}) - \log P_\theta(s_t | z_t) \right] + D_{KL}[Q_\phi \| P(Z)]$$

The expectation is estimated using a one-sample estimate and a reparametrization trick widely used in variational auto-encoders:

$$L_{consistency}(D_i; \phi, \psi) = \sum_{j=1}^K D_{KL}[Q_\psi^j \| Q_\phi];$$

$$L_{stability}(D_i; \phi, \psi) = \text{Gersh}(A_t(z_t)) + \text{Gersh}(B_t(z_t)).$$

Here  $\text{Gersh}(X)$  denotes **Gershgorin loss** [15, 16]:

$$\text{Gersh}(X) = \sum_{i=1}^n \max(0, X_{i,i} + \sum_{j \neq i} |X_{i,j}| + \varepsilon),$$

where  $|X_{i,j}|$  denotes a minor of a matrix  $X$ , and  $\varepsilon > 0$  is a small constant.

According to the Theorem 1 from [15], if the loss value is non-positive, all eigenvalues of a matrix  $X$  are guaranteed to have a negative real part, thus ensuring dynamical system stability. The usage of Gershgorin loss in composite loss function is mandatory, as ms-E2C( $K$ ) diverges for larger  $K$ .

### Algorithm

Now, we summarize an algorithm for fitting the instance of ms-E2C model we described earlier.

1. Sample a dataset of sub-trajectories using a pretrained or random policy:

$$D = \{(s_t, a_t, s_{t+1}, a_{t+1}, s_{t+2}, \dots, a_{t+K-1}, s_{t+K})_i \mid i = 1, \dots, N\}.$$

2. Initialize the weights of neural nets  $\phi, \psi, \theta$ .
3. Repeat for  $c$  epochs:
  - a) Retrieve a sample  $D_i$  from the dataset  $D$



b) Compute updated weights using a stochastic gradient descent step:

$$\phi' = \phi - \gamma \nabla_{\phi} (L_{upper}(D_i) + \lambda_1 L_{consistency}(D_i) + \lambda_2 L_{stability}(D_i));$$

$$\psi' = \psi - \gamma \nabla_{\psi} (L_{upper}(D_i) + \lambda_1 L_{consistency}(D_i) + \lambda_2 L_{stability}(D_i));$$

$$\theta' = \theta - \gamma \nabla_{\theta} L_{upper}(D_i);$$

c) Update neural networks' parameters:

$$\phi = \phi', \quad \psi = \psi', \quad \theta = \theta'.$$

Here  $\lambda_1, \lambda_2$  are tunable hyperparameters.

One might notice that unlike [15] we do not introduce an inner optimization loop to ensure stability of the internal latent space dynamics. Instead, we add the stability loss to the composed loss function. We found that although the difference is apparent during a few first epochs, it becomes negligible after a while. Stability condition does not get violated and the general results are almost the same.

## EXPERIMENTAL VALIDATION

### Planar system

Following [12–14], we use a Planar benchmark to compare the performance of the algorithms. In it, a state space is represented as a black-and-white image of a grid world with obstacles. In order to collect a dataset, we sample a random initial state and perform a series of random actions to obtain a trajectory of length  $K$ .

As in [12, 13], we use a deconvolutional network architecture [17] for image reconstruction from the latent state. For the sake of comparability, we chose the same architecture as in other papers on the topic.

The visualizations of the obtained latent state spaces are provided on a Fig. 2. The numerical results are summarized in a Table.

### Comparison of reconstruction and prediction losses

Method	State Loss $\log P_{\theta}(s_t   z_t)$	Next State Loss $\log P(s_{t+1}   s_t, a_t)$
Non-linear E2C	$9.2 \pm 4.5$	$11.7 \pm 8.8$
Global E2C	$7.6 \pm 5.7$	$10.6 \pm 5.2$
E2C	$7.6 \pm 2.3$	$10.1 \pm 2.7$
ms-E2C(3)	<u><math>7.3 \pm 1.7</math></u>	$8.7 \pm 1.9$
ms-E2C(5)	$7.6 \pm 2.1$	$7.5 \pm 1.6$
ms-E2C(7)	$7.7 \pm 2.0$	<u><math>6.3 \pm 0.9</math></u>

State loss is a regular reconstruction loss. As we observe, ms-E2C( $K$ ) give only slight average improvements on it, which is entirely expected. The introduced method does not change the architecture of a decoding network nor does it add any improvements to the algorithm regarding this matter. An important thing to notice though, is that our generalization does not make the reconstruction per-

formance much worse, which might be expected as representation is influenced by addition prediction constraints. Next state is computed by encoding the state

$s_t \xrightarrow{Q_\phi} z_t$ , predicting the next latent state  $z_t \xrightarrow{Q_\psi^1} z_{t+1}$ , and decoding the predicted regular state  $z_{t+1} \xrightarrow{P_\theta} s_{t+1}$ . Now, the results for previous methods were reproduced with slight perturbations, as we used our own codebase for it.

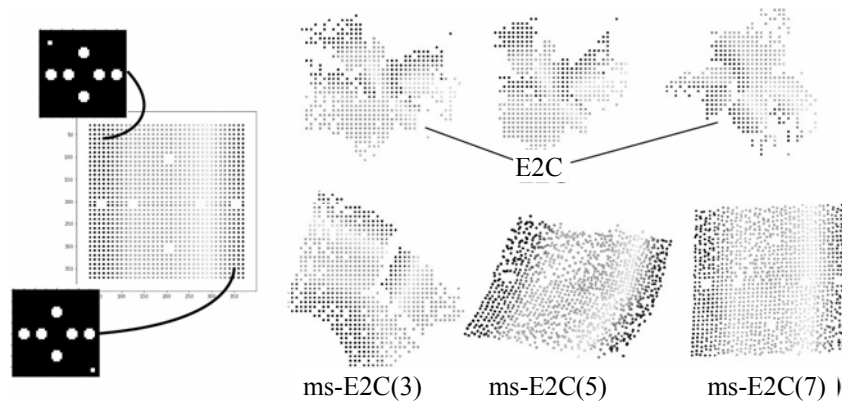


Fig 2. A comparison of latent state spaces learned by E2C and ms-E2C methods.

It's worth noting that E2C results coincide with other papers which involved reproduction of E2C [13, 14], while the original paper provides better visuals. A visualization is obtained by transforming all possible environment states with the network  $Q_\phi$ . See the scheme on the left for details.

## CONCLUSION

In this paper, a novel method had been derived as a generalization over the previous works on LCEs. We demonstrate, how the method improves upon E2C without drastic model changes which come with other works, such as PCC and P3C. We empirically show, that by considering the multistep prediction ms-E2C allows to learn a much better latent state spaces in terms of curvature and predictability, by adding a simple yet efficient way to explicitly control the desired curvature of a resulting space. At implementation is available at [18].

Moreover, our work introduces a new dimension to the LCE family of algorithms. Our future work will focus on using the approaches from the state of the art LCE methods, like predictive coding to make LCEs applicable to the higher dimensional real-world MDPs with limited amount of data to learn dynamics embedding from. We will also explore an intriguing possibility to not only encode the state, but also the action space, which sometimes has the complex structure. Lastly, we would like to study various extensions of the method to imitation learning and model-based reinforcement learning.

## REFERENCES

1. G. Dulac-Arnold et al., "Deep reinforcement learning in large discrete action spaces," *arXiv preprint arXiv:1512.07679*, 2015. doi: 10.48550/arXiv.1512.07679.

2. S. Levine, A. Kumar, G. Tucker, and J. Fu, “Offline reinforcement learning: Tutorial, review, and perspectives on open problems,” *arXiv preprint arXiv:2005.01643*, 2020. doi: 10.48550/arXiv.2005.01643.
3. E.A. Feinberg, P.O. Kasyanov, and M.Z. Zgurovsky, “Partially observable total-cost markov decision processes with weakly continuous transition probabilities,” *Mathematics of Operations Research*, vol. 41, no. 2, pp. 656–681, 2016. doi: 10.1287/moor.2015.0746.
4. E.A. Feinberg, P.O. Kasyanov, and M.Z. Zgurovsky, “Convergence of probability measures and markov decision models with incomplete information,” *Proceedings of the Steklov Institute of Mathematics*, vol. 287, no. 1, pp. 96–117, 2014. doi: 10.1134/S0081543814080069.
5. O. Vinyals et al., “Grandmaster level in starcraft ii using multi-agent reinforcement learning,” *Nature*, vol. 575, no. 7782, pp. 350–354, 2019. doi: 10.1038/s41586-019-1724-z.
6. S. Reed et al., “A generalist agent,” *arXiv preprint arXiv:2205.06175*, 2022.
7. D. Ha and J. Schmidhuber, “World models,” *arXiv preprint arXiv:1803.10122*, 2018. doi: 10.48550/arXiv.1803.10122.
8. T.M. Moerland, J. Broekens, and C.M. Jonker, “Model-based reinforcement learning: A survey,” *arXiv preprint arXiv:2006.16712*, 2020. doi: 10.48550/arXiv.2006.16712.
9. D. Hafner et al., “Learning latent dynamics for planning from pixels,” in *International conference on machine learning*, PMLR, 2019, pp. 2555–2565. doi: 10.48550/arXiv.1811.04551.
10. R.F. Prudencio, M.R. Maximo, and E.L. Colombini, “A survey on offline reinforcement learning: Taxonomy, review, and open problems,” *arXiv preprint arXiv:2203.01387*, 2022. doi: 10.48550/arXiv.2203.01387.
11. W. Li and E. Todorov, “Iterative linear quadratic regulator design for nonlinear biological movement systems,” in *ICINCO (1)*, Citeseer, 2004, pp. 222–229. doi:10.5220/0001143902220229.
12. M. Watter, J. Springenberg, J. Boedecker, and M. Riedmiller, “Embed to control: A locally linear latent dynamics model for control from raw images,” *Advances in neural information processing systems*, vol. 28, 2015.
13. N. Levine, Y. Chow, R. Shu, A. Li, M. Ghavamzadeh, and H. Bui, “Prediction, consistency, curvature: Representation learning for locallylinear control,” *arXiv preprint arXiv:1909.01506*, 2019.
14. R. Shu et al., “Predictive coding for locally-linear control,” in *International Conference on Machine Learning*, PMLR, 2020, pp. 8862–8871. doi: 10.5555/3524938.3525760.
15. M. Lechner, R. Hasani, D. Rus, and R. Grosu, “Gershgorin loss stabilizes the recurrent neural network compartment of an end-to-end robot learning scheme,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020, pp. 5446–5452. doi: 10.1109/ICRA40945.2020.9196608.
16. R.A. Horn and C.R. Johnson, *Matrix analysis*. Cambridge university press, 2012. doi: 10.5555/2422911.
17. M.D. Zeiler, D. Krishnan, G.W. Taylor, and R. Fergus, “Deconvolutional networks,” in *2010 IEEE Computer Society Conference on computer vision and pattern recognition*, pp. 2528–2535. doi: 10.1109/CVPR.2010.5539957
18. A. Tytarenko, *RL-research*. Available: <https://github.com/titardrew/rl-research>, 2022.

Received 31.08.2022

### INFORMATION ON THE ARTICLE

**Andrii M. Tytarenko**, ORCID: 0000-0002-8265-642X, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: titarenkoan@gmail.com

#### **БАГАТОКРОКОВЕ ПРОГНОЗУВАННЯ В ЛІНЕАРИЗОВАНИХ ЛАТЕНТНИХ ПРОСТОРАХ ДЛЯ НАВЧАННЯ РЕПРЕЗЕНТАЦІЙ / А.М. Титаренко**

**Анотація.** Запропоновано новий метод, що узагальнює підходи LSE, такі як E2C. Метод розвиває ідею вивчення локально-лінійного простору станів шляхом розглядання багатокрокового прогнозування, що дає змогу чіткіше контролювати кривизну шуканого простору. Продемонстровано, що метод перевершує E2C без суттєвих змін загальної моделі, на відміну від інших робіт, таких як РСС і РЗС. Розглянуто зв'язок між E2C і запропонованим методом та між їх відповідними рівняннями оновлень. Подано емпіричні докази, які свідчать, що ms-E2C дозволяє набагато краще вивчати простори прихованих станів з точки зору кривизни та прогнозованості наступних станів. Крім того, висвітлено певні проблеми стабільності, пов'язані з багатокроковими прогнозами, та способи їх вирішення.

**Ключові слова:** навчання репрезентацій, навчання керованих просторів, навчання з підкріпленням, латентний простір станів.

## ВІДОМОСТІ ПРО АВТОРІВ

**Астраханцев Андрій Анатолійович,**

доцент, кандидат технічних наук, доцент кафедри інформаційно-комунікаційних технологій та систем КПІ ім. Ігоря Сікорського, Україна, Київ

**Беспалов Вячеслав Петрович,**

науковий співробітник відділу інформатики навколишнього середовища Інституту проблем математичних машин і систем Національної академії наук України, Київ

**Бідюк Петро Іванович,**

професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

**Згуровський Михайло Захарович,**

академік НАН України, професор, доктор технічних наук, ректор КПІ ім. Ігоря Сікорського, Україна, Київ

**Зубрецька Ірина Сергіївна,**

спеціаліст з інформаційних технологій «Трейд-Сіті», Україна, Київ

**Ковалець Іван Васильович,**

старший науковий співробітник, доктор технічних наук, завідувач відділу інформатики навколишнього середовища Інституту проблем математичних машин і систем Національної академії наук України, Київ

**Кручинін Сергій Павлович,**

професор, доктор фізико-математичних наук, провідний науковий співробітник Інституту теоретичної фізики ім. М.М. Боголюбова НАН України, Київ

**Ляшенко Галина Євгенівна,**

асистент кафедри інформаційно-мережної інженерії Харківського національного університету радіоелектроніки

**Майстренко Світлана Яківна,**

старший науковий співробітник, кандидат технічних наук, старший науковий співробітник відділу інформатики навколишнього середовища Інституту проблем математичних машин і систем Національної академії наук України, Київ

**Мілявський Юрій Леонідович,**

доктор технічних наук, доцент кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

**Пишнограсв Іван Олександрович,**

доцент, кандидат фізико-математичних наук, доцент кафедри штучного інтелекту КПІ ім. Ігоря Сікорського, Україна, Київ

**Романенко Віктор Демидович,**

професор, доктор технічних наук, заступник директора з науково-педагогічної роботи ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

**Титаренко Андрій Миколайович,**

аспірант ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

**Удовенко Олег Ігорович,**

старший науковий співробітник Українського центру екологічних та водних проектів, Україна, Київ

**Федін Сергій Сергійович,**

професор, доктор технічних наук, професор кафедри інформаційних систем і технологій Національного транспортного університету, Україна, Київ

**Яценко Віталій Олексійович,**

професор, доктор технічних наук, професор факультету прикладної математики КПІ ім. Ігоря Сікорського, Україна, Київ

**G. Vijayasekaran,**

Research Scholar, Assistant Professor, Department of Computer Science and Engineering, Sir Issac Newton College of Engineering and Technology, India, Nagapattinam

**M. Duraipandian,**

Professor, Department of Computer Science and Engineering, Hindusthan Institute of Technology, India, Coimbatore

**Hemlata Goyal,**

Ph.D., Department of Computer and Communication Engineering, Manipal University Jaipur, (Raj.), India

**Sandeep Kumar Bothra,**

Research Scholar, Department of Computer Science and Engineering, Manipal University Jaipur, (Raj.), India

**Sunita Singhal,**

Associate Professor, Department of Computer Science and Engineering, Manipal University Jaipur, (Raj.), India