

Research Article

Multi-criteria Decision-Making Approach for choosing e-Bus for Urban Public Transport in the City of Niš

Nikola Petrović^{1,*}, Vesna Jovanović¹, Saša Marković¹, Dragan Marinković^{2,3}, Boban Nikolić¹

¹ University of Niš, Faculty of Mechanical Engineering
Aleksandra Medvedeva 14, 18106, Niš, Serbia

² Department of Structural Analysis, Technical University of Berlin, 10623 Berlin, Germany

³ Institute of Mechanical Science, Vilnius Gediminas Technical University, LT-10105 Vilnius, Lithuania

*e-mail: petrovic.nikola@masfak.ni.ac.rs

Submitted: 18/10/2024 Revised: 12/11/2024 Accepted: 12/12/2024 Published online: 26/02/2025

Abstract: Public urban passenger transport is a large consumer of fossil fuels, which contributes to the emission of harmful gases. Although alternative fuels are more environmentally friendly, their widespread use is still challenging even for developed countries, partly due to their higher cost. There is also some resistance to introducing sustainable transport systems due to the perception that it could significantly change the style and quality of life. Today, the bus subsystem of public passenger transport is the most widely used public transport technology. Many European public transport systems rely heavily on buses powered by conventional fossil fuel, are integral to their local fleets in most EU states. For public urban transport to increase its participation in the modal distribution of trips in cities, public transport vehicles must comply with established European standards that define different limits of exhaust gas emissions. Public transportation of passengers by bus plays a significant role in the transport system of the City of Niš as an accessible and acceptable means of transportation. Buses that use electricity as propulsion (e-Bus) are considered the cleanest technologies, producing zero local emissions and having the greatest impact on increasing local air quality. In the paper, a multi-criteria ranking of six electric buses was performed based on four criteria, available on the market of the Republic of Serbia, using the MABAC and MOORA methods, while the Entropy method was used to calculate the weight coefficients, to select an adequate bus manufacturer for the needs of public passenger transportation in the City of Niš.

Keywords: e-Bus; City of Niš; MABAC; MOORA; Entropy

I. INTRODUCTION

The EU's commitments to reduce GHG emissions are translated into concrete goals at the individual level of each member state and for individual economic sectors [1]. The White Paper [2] of the European Commission sets goals related to the reduction of GHG emissions by 60% in 2050 compared to 1990. Since around 70% of GHG emissions caused by traffic come from road traffic, the White Paper sets a target of reducing emissions by around 60%. In addition to the above, the White Paper also states the following goals: to reduce the use of "conventional fuels" in city traffic by 50% by 2030; and complete replacement of "conventional fuels" in cities by 2050 [2].

Every year, for urban areas, air quality continues to be a priority to global warming, which primarily means reducing pollutant emissions [1, 3]. European cities will face new challenges in making economically and environmentally acceptable decisions [4] as urban mobility is predicted to increase by 100% [5] and the EU targets to increase the share of public transport, as well as complying with new CO₂ emission regulations for vehicles. Achieving these goals will not only require technology that makes motorized vehicles more energy efficient but also the transition to low (or even lower) carbon dioxide emission modes of transport, such as public urban passenger transport, and motivating city dwellers to use certain modes of transport, replaced by walking and cycling.

The Low-emission Mobility Strategy adopted by the European Commission identifies key levers in the field of transport, including EU-level measures to increase the participation of low-emission and zero-emission vehicles, as well as vehicles with low-emission alternative fuels [6].

For public urban transport to increase its participation in the modal distribution of trips in cities, it is necessary to implement the improvement and development of a better network of lines to satisfy the daily need for user mobility. This would mean that the use of public urban transport should be seen as a competitor to cars. Public urban transport should be more frequent, cheaper, more reliable, safer, and accessible to all users to become more popular and provide the same or even better mobility characteristics than other forms of transport. However, it is also necessary for public transport vehicles to comply with established European standards that define different limits of exhaust gas emissions from road vehicle engines. Today, the most widely used public transportation technology is the bus subsystem of public passenger transportation. Given that buses powered by conventional fossil (diesel) fuel are large emitters of pollutants, the introduction of "cleaner" buses into daily use can contribute to reducing exhaust gas emissions and improving air quality.

In City of Niš, traffic represents one of the primary and fastest growing activities of human activities that release harmful substances or pollutants as a result of burning fossil fuels, and road traffic is the most responsible for the increase in exhaust gas emissions. The city is facing problems imposed by the large increase in the number of individual vehicles, which require more traffic areas than the city can provide and worsen living conditions and quality of life. The general importance of bus traffic for the city of Nis and its use is primarily related to the issue of urban mobility [7]. The increase in the number of vehicles and the volume of traffic threatens the quality of the environment the most, i.e., exposure of the population to poor air quality and the negative impact of noise.

After the introductory part that emphasizes the importance and analysis of the public transport system with electric drive, the second part of the paper presents an extensive review of the literature on the application of approaches and methodologies for the use of multi-criteria decision-making

techniques in public passenger transport. The third part describes the general procedure of using multi-criteria methods and defining input data that are presented as alternatives and criteria, which ultimately leads to the formation of a decision matrix. The selection of appropriate alternatives and criteria was made based on a review of foreign scientific and technical literature and the available market in the Republic of Serbia. In the fourth part, the process of applying Entropy for determining weight coefficients, as well as the use of MABAC and MOORA methods for ranking electric buses in the City of Niš, is explained. The paper will conclude with the main findings and an overview of future research objectives.

II. RELATED PAPERS

Buses are a key component of public transport, providing a cost-effective and flexible service with benefits in terms of capacity and speed. However, rising car traffic and increasing CO₂ emissions in urban areas pose significant risks to city life. To address these issues, there is a growing need for improved public transportation systems that reduce traffic congestion and utilize cleaner technologies to enhance air quality. Electric buses (EBs) are considered a crucial solution for improving urban air quality and enhancing residents' quality of life.

Public bus operators worldwide, from Shenzhen to Philadelphia and Izmir to Delhi, are increasingly adopting electric buses [8]. Their decision is driven not only by environmental concerns, such as reducing noise pollution and supporting the green transition but also by the economic advantages that emerge when evaluating the entire life cycle cost. The paper emphasizes that integrating electric vehicles requires a comprehensive approach and has the potential to revolutionize operations. Transport companies could evolve into community service providers, occasionally offering balancing energy to power supply systems through Vehicle-to-Grid (V2G) technology or acting as virtual power plants in collaboration with solar power operators. Volánbusz Zrt. has started building a data-driven ecosystem to support this model, enabling cost-optimized operations using data from its growing electric bus fleet and Industry 4.0 technologies [8].

Decision-makers face difficulties in choosing the most suitable EB due to the wide range of available options fueled by technological advancements. Multi-criteria decision-making (MCDM) methods

offer a structured approach to solving this problem. In this study [9], five electric buses are assessed using six criteria through two MCDM methods: the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and Multi-Objective Optimization based on Ratio Analysis (MOORA). These methods help rank alternatives in complex decision-making situations. The study concludes that the E₅-Bus is the best option according to both methods, with results from MOORA and TOPSIS being closely aligned. Additionally, the MOORA method is highlighted as an effective tool for solving vehicle selection problems in transportation. The proposed model has been validated with real-world applications and can support decision-makers in selecting electric vehicles for sustainable transportation [9].

Electric buses, for example, produce zero tailpipe emissions, contributing to cleaner air and making them especially desirable in densely populated areas to enhance air quality and urban livability. The paper [10] proposes a multicriteria decision-making process using the Analytic Hierarchy Process (AHP) and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) to evaluate electric bus options for central Ankara. Six electric bus alternatives were assessed based on seven specific criteria. Additionally, sensitivity analysis confirmed the robustness of these results across different scenarios [10].

The energy consumption patterns of conventional fuel-powered and electric vehicles differ due to their unique driving characteristics, a topic widely studied but often in less common geographical settings. The paper [11] fills that gap by analyzing driving data from both electric and diesel buses on the same routes in Hong Kong during regular daily operations. This allowed for a fair comparison of driving behaviors between the two bus types under identical real-world conditions, highlighting the novelty and contribution of this research. The results showed that route-specific comparisons revealed significant differences in driving patterns between electric and diesel buses, which may have been overlooked in mixed-route analyses. These differences were more pronounced in terms of range, intensity, and direction when analyzed on a route-by-route basis, affecting energy consumption. The study suggests that government agencies and bus operators should consider these findings when planning the deployment of electric buses [11].

The paper [12] focuses on the Republic of Serbia, where EVs account for only a small fraction of registered vehicles (0.007%). It also analyzes the attitudes and preferences of the Serbian population regarding EVs through a survey, identifying the key reasons for purchasing EVs and the main obstacles to adoption. The study provides useful insights for policymakers in similar markets on how to boost EV adoption and helps manufacturers understand which features are most attractive to potential buyers in these regions [12].

In the pursuit of sustainable urban development, implementing cleaner propulsion systems in public transportation is essential for reducing urban pollution and emissions. The study [7] examines the City of Niš, where traditional buses significantly contribute to environmental degradation. The necessity for alternative propulsion systems is clear, but the transition presents various challenges and uncertainties. To navigate this complexity, the research employs the CRiteria Importance Through Intercriteria Correlation (CRITIC) method to establish weight coefficients and the Evaluation based on Distance from Average Solution (EDAS) method to determine optimal propulsion systems. These methodologies enable a comprehensive evaluation of options, including buses, electric trolleybuses, and trams, for both urban and suburban transport. The study offers a systematic analysis of each alternative based on established criteria, aiding in the identification of the most effective propulsion systems. This approach not only facilitates informed decision-making aligned with sustainability objectives but also significantly mitigates the environmental impact of urban transportation. The findings provide a foundational framework for decision-makers to strategically adopt eco-friendly transport solutions in urban contexts [7].

III. MCDM METHODOLOGY, INPUT AND OUTPUT DATA

A decision is the result of a process of evaluation and choice between alternatives, with the aim to achieve a certain result. It can be strategic, tactical, or operational, and is classified by the nature of the data and the way it is delivered (intuitive, programmed, or analytical). Decisions are made under conditions of certainty, risk, or uncertainty, and can be individual, group, or collective. The quality of the decision depends on the sufficiency of information, time, complexity, and costs. In addition, the context in which the decision is made,

as well as its importance, plays a key role in the decision-making process [13, 14].

The nature of each criterion is established, specifying whether it should be minimized or maximized in the process of selecting an alternative [3]. Afterward, the alternatives are assessed for each criterion based on clearly defined parameters or subjective evaluations.

Electric buses represent a relatively new technology that is constantly being improved, especially in the area of energy storage systems (batteries), vehicle charging, traction control, optimization of energy consumption, and reduction of empty vehicle mass. Today, almost all the world's bus manufacturers offer electric buses of various sizes, including midi (8-9 meters), standard (11-13 meters), articulated (18-19 meters) and double-articulated (24-27 meters). Also, e-Buses have become part of the standard offer on the market. According to the ZEUS report [15], 32 manufacturers of electric buses and 8 manufacturers of charging systems are registered on the European market. The most represented manufacturers on the European market include BYD, VDL, Solaris, Volvo, Kamaz, GAZ, Yutong, Ebusco, Optare, Caetano, Skoda, Irizar and Van Hool.

When ranking electric buses for public passenger transport in the City of Niš, six alternatives (electric bus manufacturers) are being explored from A_1 to A_6 , taking into account the latest technological advancements that are available on the market in the Republic of Serbia. The evaluation is based on several criteria:

C_1 - Price (in thousands of \$) - The price of city electric buses depends on various factors such as bus size, battery capacity, manufacturer, and additional technology used. Although electric buses are more expensive to purchase, maintenance and fuel costs are significantly lower than conventional buses, which can lead to savings over a longer time.

C_2 - Charging time (in hours) - Represents the time required to charge the battery from the designed minimum to the maximum value of the battery capacity. Batteries are the most expensive components on an electric bus. Reducing the costs of these systems and improving their operating characteristics are of key importance for even greater applications in electric buses.

C_3 - Total number of seats (standing and seated) The total number of seats in the bus depends on the

manufacturer, and model of the bus, its purpose and design, as well as the regulations of the country in which it is used.

C_4 - Range (in km maximum) - The range of city electric buses varies depending on the model, battery capacity, and the conditions in which they are used. Factors such as terrain, traffic density, air conditioning, and bus load can affect the actual range. Many electric buses use energy regeneration systems, which can extend the range during city driving.

A decision matrix (**Table 1**) has been created using the selected alternatives and criteria. This matrix helps determine the weight coefficients assigned to each criterion, which affect the selection of alternatives. Following this, a multi-criteria decision-making process is carried out, utilizing the determined weight coefficients to evaluate and rank the alternatives efficiently.

Table 1. The initial decision matrix

	C_1	C_2	C_3	C_4
	<i>min</i>	<i>min</i>	<i>max</i>	<i>max</i>
A_1	1120	1.4	83	180
A_2	750	3.1	79	300
A_3	1000	3.2	95	320
A_4	592.6	2.5	62	253
A_5	720	3.6	90	350
A_6	842	4	70	400

IV. IMPLEMENTATION OF THE APPROACH AND DISCUSSION OF THE RESULTS

MCDM tools are commonly applied to address complex decision problems, particularly in transportation. One important decision area involves selecting clean technology vehicles, such as electric buses, which offer distinct advantages over traditional internal combustion engine vehicles.

1. Entropy method

Determining the objective weights of the criteria using the entropy method is based on the measurement of the degree of uncertainty of the information contained in the decision matrix. This approach directly generates criteria weight values based on the contrast between the values of the alternatives for each criterion, considering all criteria simultaneously. It is considered an objective method because the weights derive from the criteria values themselves, eliminating the subjectivity, incompetence, or absence of the decision maker

[16]. The nature and type of criteria are not of decisive importance for the application of this method. In the context of multi-criteria decision-making, entropy is used to determine how significant the difference is between the values of

alternatives concerning to each criterion. If the values within one criterion are very similar, then the information provided by that criterion is low, and vice versa. Algorithm steps are [17, 18]:

In the first step, a decision matrix is formed:

$$X = \begin{matrix} & C_1 & C_2 & \dots & C_n \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{matrix} & \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \end{matrix} \quad (1)$$

In the second step, all members of the matrix x_{ij} are normalized using the form:

$$r_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}} \quad (2)$$

In the following third step, the amount of information contained in the normalized decision matrix and emitted by each criterion f_j can be measured as the entropy value e_j :

$$e_j = -k \sum_{i=1}^m r_{ij} \ln r_{ij} \quad j = 1, 2, \dots, n \quad (3)$$

By introducing the constant $k=1/\ln m$, it is ensured that all e_j values are in the interval $[0, 1]$.

In the next step, the degree of divergence d_j is determined in relation to the average amount of information contained in each criterion. Since the value d_j represents a specific measure of the contrast intensity of the criterion f_j , the final relative weight of the criterion can be obtained by simple additive normalization (**Table 2**):

$$w_j = \frac{d_j}{\sum_{j=1}^n d_j} \quad (4)$$

Table 2. Obtained weighting coefficients of criteria

	C_1	C_2	C_3	C_4
w_j	0.2101	0.4221	0.0959	0.2718

The Entropy method is useful because it removes the subjectivity in determining the weights and allows the decision-maker to rely on the mathematical characteristics of the data itself.

The calculation results of weighting coefficients, obtained through software MS Excel based on Eq (1-4), that the C_2 -Charging time (0.4221) is the most important criterion in the evaluation system. After that comes a criterion C_4 -Range (0.2718) and C_1 -Price (0.2101). Criterion C_3 -Total number of seats (standing and seated) has the smallest weighting coefficient (0.0959) showing that this indicator has a minor impact on the evaluation process.

2. MABAC (Multi-Attributive Border Approximation Area Comparison) method

It is one of the newer methods for solving multi-criteria decision-making (MCDM) problems. Its main feature is the introduction of the concept of a marginal approximation region for each alternative, where the performance of the alternatives is compared to the threshold values of the criteria. The MABAC method combines the advantages of other methods, such as simplicity of application and providing precise results. The steps of applying the MABAC method are [19]:

1. Formation of the decision matrix:

A decision matrix is set up in which the rows are alternatives and the columns are criteria. The values within the matrix represent the performance of each alternative against each criterion.

2. Normalization of values

The values in the matrix are normalized to bring them to the same scale. The following formula is used for the criteria to be maximized:

$$r_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (5)$$

The criteria to be minimized:

$$r_{ij} = \frac{\max(x_j) - x_{ij}}{\max(x_j) - \min(x_j)} \quad (6)$$

Here, r_{ij} is the normalized value of alternative i for criterion j .

3. Calculation of the boundary value matrix (BV):

The threshold values for each criterion represent the average of the normalized performance values for that criterion according to the following formula:

$$GB_j = \frac{1}{m} \sum_{i=1}^m r_{ij} \quad (7)$$

where BV_j is the threshold value for criterion j , and m is the number of alternatives.

4. Calculation of the distance matrix of the alternative from the threshold values according to the formula:

$$q_{ij} = r_{ij} - BV_j \quad (8)$$

where q_{ij} is the distance of the normalized value of alternative i for criterion j from the threshold value BV_j .

5. Calculating the approximation function G :

For each alternative, the approximation function is calculated as the sum of weight-corrected values q_{ij} :

$$G_i = \sum_{j=1}^n w_j \cdot q_{ij} \quad (9)$$

where G_i is the approximation function value for alternatives i , w_j is the weight of criterion j , and q_{ij} is the distance value for alternatives i .

6. Alternative ranking. The alternatives are ranked based on the calculated G_i values. An alternative with a higher G_i value is considered better and is ranked higher.

The MABAC method allows the decision maker to see how each alternative is positioned to the "boundary" that represents the average performance values for each criterion, which makes this method very effective for solving multi-criteria decision-making problems.

3. MOORA method

It is simple to use and provides effective results in solving multi-criteria decision-making problems.

The algorithm of applying the MOORA method is as follows [20]:

After the first step, that is, the formation of the decision-making matrix, the decision-making matrix is normalized as a second step, according to the formula:

$$x_{ij}^* = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}} \quad (10)$$

In the third step, the normalized performance is added in the case of maximization and subtracted in the case of minimization, so that the optimization problem is solved according to the formula:

$$y_i = \sum_{j=1}^g w_j x_{ij}^* - \sum_{j=g+1}^n w_j x_{ij}^* \quad (11)$$

The resulting value of y_i can be positive or negative based on the total value of the maximization terms and the minimization terms. The best-ranked alternative has the highest y_i value, and the worst-ranked alternative has the lowest value.

According to the results (**Table 3**) derived from the Entropy method for evaluating weight coefficients and MABAC using Eq. (5-9), and MOORA using Eq. (10-11) methods for ranking e-Buses, in the MS Excel, the best alternative identified is A₄ manufacturer of an electric bus whose criteria are: C₁=592600 \$ (price), C₂=2.5 h (charging time), C₃=62 (Total number of seats - standing and seated), and C₄=253 km (range). The worst alternative identified is A₆ whose criteria are: C₁=842000 \$ (price), C₂=4 h (charging time), C₃=70 (Total number of seats - standing and seated), and C₄=400 km (range).

Table 3. Ranking of alternatives according to criteria

	C_1	C_2	C_3	C_4	<i>Entropy - MABAC</i>		<i>Entropy - MOORA</i>	
	<i>min</i>	<i>min</i>	<i>max</i>	<i>max</i>	G_i	Ranking	ξ_i	Ranking
w_j	0.2101	0.4221	0.0959	0.2718				
A ₁	1120	1.4	83	180	1.47614	4	-0.08550	2
A ₂	750	3.1	79	300	1.48421	3	-0.10223	3
A ₃	1000	3.2	95	320	1.43956	5	-0.11792	5
A ₄	592.6	2.5	62	253	1.53688	1	-0.07807	1
A ₅	720	3.6	90	350	1.50873	2	-0.10385	4
A ₆	842	4	70	400	1.39885	6	-0.13017	6

V. CONCLUSION

The ever-increasing competitiveness and complexity of the market, and the rapid development of technique and technology have made the decision-making process in many institutions, companies, and institutions of key, strategic importance, bearing in mind the fact that it represents one of the most important processes that takes place and has far-reaching consequences on its success and market position. The complexity of the nature of the process often imposes the need to make multi-criteria decisions. Knowing the basic theoretical concepts and the essence of multi-criteria analysis enables managers, engineers, planners and all other decision-makers to effectively apply the methods of multi-criteria analysis to solve numerous decision-making problems at different levels. With the use of computers and various software packages, multi-criteria analysis problems can be solved in a relatively short time.

Based on the shown trends of the increase in the number of electric-powered buses in operation worldwide, the large number of tender procedures around the world related to the purchase of electric-powered buses, the ever-increasing offer on the market by almost all bus manufacturers, positive experiences in exploitation and legal regulations that prescribe the increasing participation of "clean" buses in the transport systems of cities, which is best seen on the example of the member states of the European Union (Directive 94/2014/EC).

The aim and task of this work was to determine the best and most optimal solution when choosing an electric bus for the City of Niš, which can be found on the market of the Republic of Serbia, and all this by applying multi-criteria decision-making methods, specifically in this case the MABAC and MOORA methods while are the weight coefficients of the criteria calculated by the Entropy method. Based on the results obtained from the comparative analysis of the ranked alternatives based on the criteria, it can be

concluded that the best-ranked alternative is A₄ and the worst alternative identified is A₆.

The findings provide valuable insights that can guide efforts to enhance air quality in the City of Niš. By utilizing these results, the City can focus on refining its traffic development strategies, both for the present and the future. These strategies can be adjusted to prioritize sustainable and efficient traffic solutions, ultimately contributing to cleaner air and a healthier urban environment.

ACKNOWLEDGEMENT

This research was financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Contract No. 451-451-03-65/2024-03).

AUTHOR CONTRIBUTIONS

N. Petrović: Conceptualization and Supervising.

V. Jovanović: Conceptualization and Experiments.

S. Marković: Writing and Reviews.

D. Marinković: Conceptualization and Supervising.

B. Nikolić: Writing and Reviews.

DISCLOSURE STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID

N. Petrović <https://orcid.org/0000-0002-9166-1263>

V. Jovanović <https://orcid.org/0000-0001-9252-7894>

S. Marković <https://orcid.org/0009-0003-6334-7415>

D. Marinković <https://orcid.org/0000-0002-3583-9434>

B. Nikolić <https://orcid.org/0000-0002-9694-6719>

REFERENCES

- [1] N. Petrović, N. Bojović et al., A Two-Phase Model for the Evaluation of Urbanization Impacts on Carbon Dioxide Emissions from Transport in the European Union, *Tehnički vjesnik* 30 (2) (2023) pp. 514–520 <https://doi.org/10.17559/TV-20221018103946>
- [2] European Commission (2011), White paper on transport: Roadmap to a single European Transport area — towards a competitive and resource-efficient transport system, Publications Office of the European Union [cited 2024-10-14] <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0144:FIN:EN:PDF>
- [3] N. Petrović, V. Jovanović et al., Multicriteria Sustainability Assessment of Transport Modes: A European Union Case Study for 2020. J.

- Green Econ. Low-Carbon Dev. 3 (1) (2024) pp. 36–44 <https://doi.org/10.56578/jgelcd030104>
- [4] European Commission, A Roadmap for moving to a competitive low carbon economy in 2050, [cited 2024-10-14] <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0112:FIN:en:PDF>
- [5] McKinsey (2012), Urban buses: alternative powertrains for Europe. A fact-based analysis of the role of diesel-hybrid, hydrogen, fuel cell, trolley and battery electric powertrains https://www.fuelcellbuses.eu/sites/default/files/documents/20121029%20Urban%20buses%20C%20alternative%20powertrains%20for%20Europe%20-%20Final%20Report_0_0.pdf
- [6] B. Nikolić, B. Kegl et al., Characteristics of biodiesel as a fuel for diesel engines, *Innovative Mechanical Engineering*, University of Niš, Faculty of Mechanical Engineering 2 (3) (2023) pp. 66–78.
- [7] N. Petrović, S. Marković et al., Evaluating Alternative Propulsion Systems for Urban Public Transport in Niš: A Multicriteria Decision-Making Approach. *J. Eng. Manag. Syst. Eng.*, 3 (2) (2024) pp.72–81 <https://doi.org/10.56578/jemse030202>
- [8] V. Kruchina, The possibility of electrification in public transport bus services, *Acta Technica Jaurinensis* 16 (4) (2023) pp. 158–166 <https://doi.org/10.14513/actatechjaur.00713>
- [9] Hamurcu, M., Eren, T. Applications of the MOORA and TOPSIS methods for decision of electric vehicles in public transportation technology, *Transport* 37 (4) (2022) pp. 251–263 <https://doi.org/10.3846/transport.2022.17783>
- [10] M. Hamurcu, T. Eren, Electric Bus Selection with Multicriteria Decision Analysis for Green Transportation, *Sustainability* 12:2777 (2020) <https://doi.org/10.3390/su12072777>
- [11] K.-W. Ng, H.-Y Tong, Comparisons of Driving Characteristics between Electric and Diesel-Powered Bus Operations along Identical Bus Routes, *Sustainability* 16 (12):4950 (2024) <https://doi.org/10.3390/su16124950>
- [12] N. Pajic, N. Jurišević et al., Adoption of Electric Vehicles in the Republic of Serbia, *Transportation in Developing Economies*, 10 (33) (2024) <https://doi.org/10.1007/s40890-024-00220-2>
- [13] N. Petrović, V. Jovanović et al., Evaluating the operation performance of the Serbian transport freight system by using Multiple Criteria Decision-Making technique, *Engineering Today* 1 (4) (2022) pp. 33–40 <https://doi.org/10.5937/engtoday2204033P>
- [14] M. Radovanović, D. Božanić et al., Application of Hybrid DIBR-FUCOM-LMAW-BONFERRONI-GREY-EDAS model in multicriteria decision-making, *Facta Universitatis-Series Mechanical Engineering*, 21 (3) (2023) pp. 387–403 <https://doi.org/10.22190/FUME230824036R>
- [15] ZeEUS, (2017) An updated overview of electric buses in Europe [cited 2024-10-14] <https://zeeus.eu/uploads/publications/documents/zeeus-ebus-report-2.pdf>
- [16] N. Petrović, J. Mihajlović et al., Evaluating Annual Operation Performance of Serbian Railway System by using Multiple Criteria Decision-Making Technique, *Acta Polytechnica Hungarica*, 20 (1) (2023) pp. 157–168 <https://doi.org/10.12700/APH.20.1.2023.20.11>
- [17] E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana: University of Illinois Press (1947)
- [18] C. N. Wang, T. Q. Le et al., Measuring road transport sustainability using MCDM-based entropy objective weighting method, *Symmetry* 14 (5) 2022 <https://doi.org/10.3390/sym14051033>
- [19] D. Božanić, D. Pamučar et al., Primene metode MABAC u podršci odlučivanju upotrebe snaga u odbrambenoj operaciji, *Tehnika–Menadžment*, 66 (2016) pp. 129–136 <https://doi.org/10.5937/tehnika1601129B>
- [20] W. Brauers, K. E. Zavadskas et al., Multi-objective decision-making for road design, *Transport* 23 (3) (2006) pp. 183–193 <https://doi.org/10.3846/1648-4142.2008.23.183-193>



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license.

Research Article

Optimization Location Routing Problem (LRP) of Humanitarian Aid Distribution Using HGASA Method in Sigi District

Serli Bombang^{1,*}, Bonivasius Prasetya Icthiarto¹, Humiras Hardi Purba¹

¹Departement of Industrial Engineering, Mercu Buana University
Jl. Raya Meruya Selatan No.01, Kembangan, Jakarta Barat 1650, Indonesia
*e-mail: Sbombang@gmail.com

Submitted: 10/12/2024 Revised: 30/01/2025 Accepted: 31/01/2025 Published online: 26/02/2025

Abstract: Disasters are events that disturb and threaten people's lives caused by nature or/and non-natural factors as well as human factors that cause casualties and economic losses. The polemic of uneven assistance and delays in the event of a natural disaster is the most common thing that occurs during a natural disaster. Location Routing Problem is a continuation of the classical routing problem that combines strategic and operational decisions with the facility location problem and the vehicle routing problem. This study aims to determine the location of the distribution centre construction and the optimal route using the Hybrid Genetic Algorithm and Simulated Annealing methods with the objective function of minimizing total costs and minimizing maximum service time for the distribution of humanitarian aid natural disaster 2018 in Sigi Regency. Optimization is designed into two scenarios, namely the construction of two distribution centres and three distribution centres. The result show that construction of two distribution centres can be designed at locations D3 and D5 with a total cost of IDR405 609 000 and a maximum travel time of 25.921 hours, while the construction of three distribution centres can be done at locations DC1, DC4 and DC 5 or with a total cost of IDR605,778,000 and a maximum service time of 19.634 hours.

Keywords: Disaster, Genetic Algorithm, Simulated Annealing, Optimization, Location Routing Problem.

I. INTRODUCTION

Disaster is an event that disrupts and threatens people's lives caused by nature or/and non-natural factors and human factors that cause loss of life and economic losses. Natural disasters are serious things that threaten the lives and security of people's property around the world caused by natural factors such as earthquakes, landslides, tsunamis, floods and others [1]. Over the past ten years, disaster data from EM-DAT [2] states that the number of natural disasters that occurred during the last ten years from 2003-2023 has increased. The types of natural disasters that occurred include drought, earthquakes, extreme weather, floods, landslides, storms, volcanic activity and fires. As a result of the disaster, the number of individuals affected by the disaster, in 2023, was 93.1 million, below the 2003-2022 annual average of 175.5 million. Economic losses, the reported figure of US\$ 202.7 billion, were slightly higher than the EM-DAT 2003-2022 annual average of US\$ 196.3 billion. A comparison of the number of natural disasters that occurred during 2023 and those that occurred during 2003-2022 based on their

type. The number of occurrences of disasters is shown in the following **Fig. 1**.

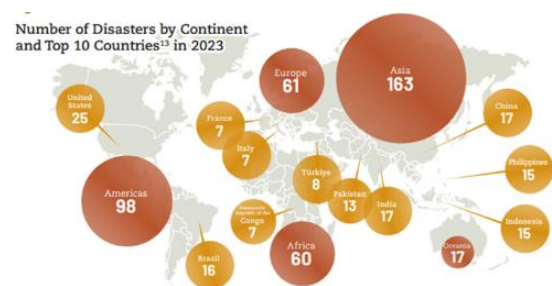


Figure 1. Number Occurrence of Disaster 2023 [2]

From **Fig. 1** shows that in Asia, Indonesia is one of the countries with a moderate frequency of natural disasters, which occurred 15 times during 2023 [2]. This shows that Indonesia is one of the countries prone to natural disasters. Central Sulawesi is one of the provinces in Sulawesi that has a relatively high Indonesian Disaster Risk Index value of 146.07[3]. Sigi Regency is one of the regencies in Central Sulawesi which consists of 15 sub-districts. Most of

its areas have a relatively high potential for disasters. The geographical conditions of Sigi Regency, which consists of the Sigi Regency area, is one of the regencies whose areas felt the direct impact of the 2018 natural disaster. According to official data from the Sigi Regency Government, 289 people died and 93,187 people had to evacuate because their homes and residential areas were damaged. Many houses were severely, moderately, or lightly damaged. In addition, various economic facilities such as markets, shops, and other accommodations, as well as public and social facilities, government offices, transportation infrastructure, communications, clean water, and other public services were also disrupted [4]. Seeing the condition of the Republic of Indonesia regarding natural disasters, the government has passed Law No. 24 [1] concerning Disaster Management. This law was issued as a form of government accountability for all Indonesian people. This law also serves as a legal basis for disaster management in Indonesia, which is known as a disaster-prone country. Logistics is one of the important elements in disaster management operations because it affects human survival or life. Logistics assistance that is right on target, at the right time, and benefits many people is the goal of humanitarian aid logistics assistance [5].

Therefore, disaster relief distribution is one of the most important components in the disaster emergency response process. All aid sent from various parties needs to be distributed quickly, accurately and evenly to victims of natural disasters. To save more lives and reduce losses caused by disasters, many researchers have conducted intensive studies on the distribution of emergency supplies [6].

The polemic of uneven assistance and delays in the event of a natural disaster is the most common thing that occurs during a natural disaster. As at the time of the 2018 earthquake, tsunami and liquefaction in Central Sulawesi, showing that until the fifth day after the earthquake, there were still many residents who did not receive logistical assistance such as food and drinks. Limited supply of food and beverages which then causes the picking of goods at stores such as minimarkets [7]. In Sigi Regency, the delay in natural disaster relief was due to the number of refugees reaching thousands of people, damage to many severely affected institutional facilities and public offices that could only operate after the fourth day after the incident and the existence of several locations that are still isolated such as Lindu, Kulawi, South Kulawi and Pipikoro after almost one month after the natural disaster[3]. Even at the beginning of the evacuation period, the community independently made simple tents from materials such as tarpaulins and from used banners/billboards.

Therefore, the distribution of disaster assistance is one of the most important components in the overall

disaster emergency response process. All aid sent from various parties needs to be distributed quickly, appropriately and evenly for victims of natural disasters. So, to save more lives and reduce losses caused by disasters, many researchers have conducted intensive studies on the distribution of emergency supplies[6]. To effectively reduce the damage caused by the earthquake, aid must be sent to the point of demand in the shortest possible time after the earthquake has occurred. Experts generally believe that the problem of locating emergency facilities (LAP) and the problem of routing emergency vehicles (VRP) in the available transportation network are two of the most challenging issues in the emergency logistics system after an earthquake [8]. The Location-Routing Problem (LRP) is an area that is developing in transportation planning research. LRP covers all three levels of decision-making in supply chain management by planning facilities, vehicles, and supply network routes simultaneously [9]. The important role of distribution centres (DCs) and delivery routes is needed in distributing and transporting aid items [10].

Location Routing Problem is a continuation of the classic routing problem that combines strategic and operational decisions with facility location problems and vehicle routing problems[11]. LRP includes a combination of determining the location of the distribution centre with the selection of the vehicle routing problem. Research on the Location Routing Problem in the distribution of natural disaster relief has been conducted by several researchers ([12], [13], [14], [15], [16], [8], [17]). Meanwhile, research for solving the optimization problem of HGASA method has been used by several researchers [18], [19], [20], [21], [22], [23]).

Through the review of the above literature, it can be found that academics at home and abroad have conducted intensive research on LRP with different disaster backgrounds, and most scholars focus on natural disasters such as earthquakes and floods. Through this study, the researcher will conduct research for the optimization of LRP for the distribution of natural disaster assistance in Sigi Regency Hybrid Genetic Algorithm and Simulated Annealing (HGASA) which is also a method that has been widely used in optimization problem but is still very rarely used in optimization LRP for aid distribution Humanity.

II. LITERATURE REVIEW

1. Distribution of Logistics Assistance

The distribution of logistical assistance is a way of distributing and/or providing logistical assistance for disaster management from the place of origin to the destination area to the intended target. Logistical assistance for disaster victims during the state of

emergency must be received by victims in need on time, in the right location, on target, in the right quantity, and with the right quality[3]. One of the main aspects that affects the success of logistics is the role of transportation in distributing humanitarian aid, both the type of mode of transportation and the supporting infrastructure has a great influence [5]. Humanitarian logistics involves a structured approach to overcoming natural disasters and technology. The main stages consist of four phases in the disaster management cycle, namely mitigation, preparation, emergency response, and recovery [14]. Then Hamzani et al. [12] explain the main challenge in the distribution of aid in large-scale disasters is the importance of making decisions quickly and minimizing delays.

2. Location Routing Problem

Location Routing Problem (LRP) is a combination of two NP-hard problems (Facility Location Problem (FLP) and Vehicle Routing Problem (VLP)). The Location Routing Problem (LRP) is a continuation of the classic routing problem that integrates strategic and operational decisions with facility location problems (FLP) and vehicle routing problems (VRP) [11]. Shen et al. [16] addressed the optimization problem in emergency logistics systems, integrates the perspective of environmental protection with a holistic optimization approach to emergency logistics systems. The fuzzy-based low-carbon open route-location problem model (FLCOLRP) in emergency logistics was developed with multi-objective objective functions, which include minimizing delivery time, total cost, and carbon emissions. The problem of location route has a crucial role in supply chain management. In this context, the decision taken involves determining the location of facilities, such as distribution centers, as well as the preparation of vehicle routes [24]. Also, Veysmoradi et al. [17] conducted LRP research for aid distribution with a focus on reducing the fixed cost of building distribution centers (DCs) and travel costs, minimizing the longest travel times, and improving route reliability for all vehicles involved in the process. Yan et al. [25] the research focuses on a two-stage cold chain logistics distribution network, including the optimization of transfer station locations and route planning in two stages. The two-stage location-route model was developed to minimize the total cost by considering the constraints of a tight time window. Various types of vehicles are considered to support distribution activities, and an integrated approach is used in the design of algorithms to guarantee the quality of the resulting solutions. Similarly, the research conducted by Heidari et al. [26] on the Two-Tier Open and Closed Routes (2E-COLRP) problem covers two layers, namely factories, depots, and customers, with a focus on reducing costs and CO2 emissions. This model is designed to identify the

best routes, the optimal number of vehicles and facilities, and the determination of the most strategic facility locations.

3. Hybrid Genetic Algorithm and Annealing Simulated (HGASA)

Genetic Algorithm (GA) has effective global search capabilities to quickly find solutions in a solution space. However, it has not been able to conduct local searches that can cause early convergence [20]. GA takes inspiration from the principles of biological evolution and applies a population-based approach by using genetic operators such as crossovers and mutations. On the other hand, simulated annealing (SA) is influenced by the annealing process, and utilizes probability methods for solution space exploration. Although both are used in optimization tasks, they have different mechanisms and advantages. GA is more suitable for global exploration, while SA is effective in avoiding local optimal points through probability measures. Hybrid genetic algorithm and Simulated Annealing (HGASA) is a method that combines GA and SA, which has excellent global search capabilities that can be obtained using GA and the ability to find local optimal using SA [18]. GA is used to get the optimal or near-optimal solution among the solution space, and then SA is used to search for a better one based on the solution. Genetic algorithms have the advantages of strong global optimization capabilities, fast speed, strong versatility, and easy implementation. However, it has the disadvantage of poor local search capabilities, which lowers search efficiency, especially in the late period of optimization. Fortunately, simulated annealing algorithms have strong local search capabilities to make up for the shortcomings of genetic [23].

4. Stopping Criteria's

Determining stopping criteria is an important step in the design of genetic algorithms. This criterion will determine when the algorithm should be stopped, either after achieving the desired result or when it meets the predetermined limits. Some commonly used termination criteria in genetic algorithms include [27]: (i) Number of Generations: The algorithm will stop after a certain number of generations have been evaluated. This is a simple approach where the limit on the number of generations is predetermined. (ii) Fitness Level: The algorithm is stopped when one of the individuals in the population reaches a certain level of fitness. For example, when an expected match is reached, the algorithm may stop. (iii) Convergence: Cessation can be done when the population experiences convergence, i.e. when the average value of match or variation in the population becomes stable. If the population reaches convergence quickly, the

algorithm may be stopped early. (iv) Execution Time: The algorithm is stopped after reaching a predetermined time limit, which is useful in situations with time constraints. (v) Rate of Change: If there is no significant change in the population match value over several generations, the algorithm can be stopped, indicating that there is no further progress. (vi) Combination Criteria: Multiple criteria can be combined, such as the number of generations along with a specific level of compatibility or time limit.

III. METHODOLOGY

The first step in this research is Designing a Mathematics Model for location routing problem, determine stopping criteria and parameters will using. Then conducting tests on the parameters used, optimize using HGASA for two scenarios namely two or three DC will be open with several simulations. From these result total cost and maximum travelling time will be compared using DIS and conclusion is obtained.

1. Sample

The sample used in this study is Sigi Regency which is one of the districts in Central Sulawesi Province in Indonesia which is classified as a high natural disaster-prone area. Sigi Regency is one of the areas affected by the 2018 natural disaster in Central Sulawesi province. The number of affected villages was 160, and a sample of 72 villages was taken which was classified as experiencing the severe impact of the 2018 natural disaster due to the large number of victims and severe economic damage. Map of the research can be seen in **Fig. 2**.

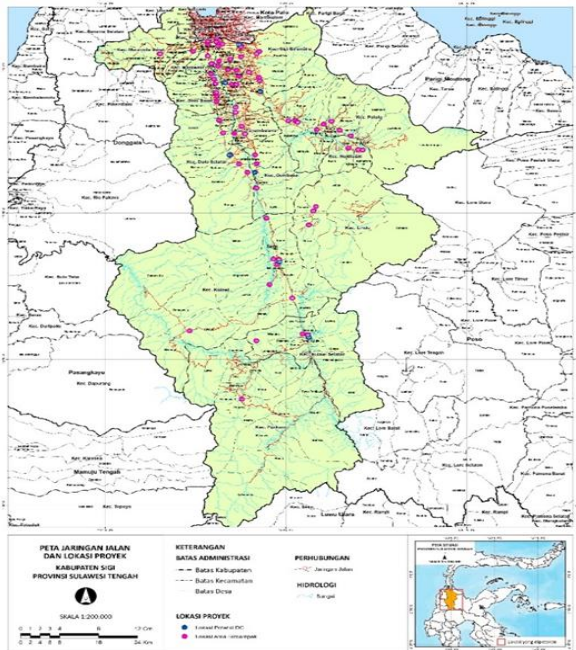


Figure 2. Sigi Regency Research Map

2. Mathematical Model

The mathematical model in this study is used to design the program in Matlab. As for some of the assumptions used in this study: (i) the number of potential distribution centers (DCs) and affected areas or distribution points are known (ii) potential locations to establish distribution centers (DCs) have been identified (iii) the shape of the route between the distribution center (DC) and the distribution point can be represented as a graph (iv) The selection of distribution centers, land routes and air routes are interrelated influenced by the constraints and function of the destination and (v) air routes can only be served by air vehicles.

Here is some consideration of the index set, parameters, dependent variables, decision variables, and objective function used in this research.

Index sets:

I : Set of potential locations of DC construction sites (1,2,3 ... i)

A : Set of potential locations affected by the disaster (1,2,3 ... j)

H : Set of available terrestrial vehicles (1,2,3 ... h)

G : Set of available aerial vehicles (1,2,3 ... g)

Parameters:

fi : investment cost of DC development at the i th location, $i \in I$

ch : variable cost of transportation per unit distance for

terrestrial vehicles h , for $h \in H$

cg : variable cost of transportation per unit distance

for aerial vehicles g , for $g \in G$

dij : The distance between the location i and the distribution point j , for $i \in I, j \in J$.

$vijh$: number of transport volume from location i to

distribution point j by terrestrial vehicle h

$uijg$: Number of transport volume from location i to

distribution point j by aerial vehicle g

vh : The speed of the terrestrial vehicle h , for $h \in H$

vg : The speed of the aerial vehicle g , for $g \in G$

kh : terrestrial vehicle capacity

kg : aerial vehicle capacity

Mj : DC capacity

Objective function:

The first objective function is for minimizing total cost, where LRP simultaneously determines the number and location of distribution centers (DCs) while allocating earthquake-affected areas to DCs and vehicle routes.

Therefore, the total cost consists of the following components: (1) fixed cost for establishing DCs and (2) cost of air and land vehicle travel per km.

$$\begin{aligned} \min f_1 \\ = \sum_{i \in I} f_i x_i + \sum_{h \in H} \sum_{(i,j) \in V} c_h d_{ij} Y_{ijh} \\ + \sum_{g \in G} \sum_{(i,j) \in V} c_g d_{ij} Z_{ijg} \end{aligned} \quad (1)$$

The second objective function minimizing maximum travel time

The objective function aims to reduce the maximum travel time of a vehicle route.

$$\begin{aligned} \min f_2 = \max \left\{ \sum_{i,j \in V} \frac{d_{ij} Y_{ijh}}{(v_{h1})}, \sum_{i,j \in V} \frac{d_{ij} Y_{ijh}}{(v_{h2})}, \right. \\ \left. \sum_{i,j \in V} \frac{d_{ij} Z_{ijg}}{(v_g)} \right\} \end{aligned} \quad (2)$$

And there are several constraints that are used in this study.

a. Vehicle Functions

$$v_{ijh} Y_{ijh} \leq K_h \quad \forall (i,j) \in V, \forall h \in H \quad (3)$$

$$u_{ijg} Z_{ijg} \leq K_g \quad \forall (i,j) \in V, \forall g \in G \quad (4)$$

b. Distribution center capacity

$$\sum_{i \in I} v_{ijh} Y_{ijh} + \sum_{i \in I} u_{ijg} Z_{ijg} \leq M_j \quad (5)$$

c. Distribution center opening restrictions

$$\sum_{i \in I} x_i \geq 1 \quad (6)$$

d. Assignment

$$\sum_{i \in I} y_{ijh} + \sum_{i \in I} z_{ijg} = 1 \quad \forall j \in J, \forall i \in I \quad (7)$$

e. Vehicle Type

$$\sum_{h \in H} y_{ijh} + \sum_{g \in G} z_{ijg} \leq 1 \quad \forall i \in I; \forall j \in J; i \neq j \quad (8)$$

f. Travel Time

$$\sum_{i,j \in V} \frac{d_{ij} Y_{ijh}}{(v_h)} \leq t_{\max, \text{terrestrial}} \quad \forall h \in H \quad (9)$$

$$\sum_{i,j \in V} \frac{d_{ij} Z_{ijg}}{(v_g)} \leq t_{\max, \text{aerial}} \quad \forall g \in G \quad (10)$$

g. Decision Variables

$$x_i \in \{0,1\} \quad \forall i \in I; \forall j \in J; \forall g \in G \quad (11)$$

$$y_{ijh} \in \{0,1\} \quad \forall (i,j) \in V; h \in H; \quad (12)$$

$$z_{ijg} \in \{0,1\} \quad \forall i \in V; g \in G \quad (13)$$

The objective function (1) for total cost minimization includes a fixed cost component for DC establishments and air and ground vehicle travel costs per km. Objective function (2) minimizing the maximum travel time of a vehicle route. Obstacle function (3-4) The number of goods transported by a vehicle does not exceed the capacity of the vehicle; (5) the transport volume of land and air vehicles does not exceed the DC capacity; (6) the number of DCs to be opened is at least one; (7) Each affected location can only be served by one DC; (8) each route is served by only one type of vehicle (land or air), and not by both; (9-10) the maximum travel time for ground and air vehicles shall comply with the established time limits; and (11-13) Decision and non-negative variables

3. Hybrid Genetic Algorithm and Annealing Simulated (HGASA)

The Hybrid Genetic Algorithm and Simulated Annealing (HGASA) is a technique that integrates the strengths of Genetic Algorithms (GA) and Simulated Annealing (SA). This approach leverages the robust global search capabilities of GA to explore the solution space for an optimal or near-optimal solution, followed by SA's proficiency in refining the solution to find a better local optimum ([18]. The stages of HGASA are shown in **Fig. 3**.

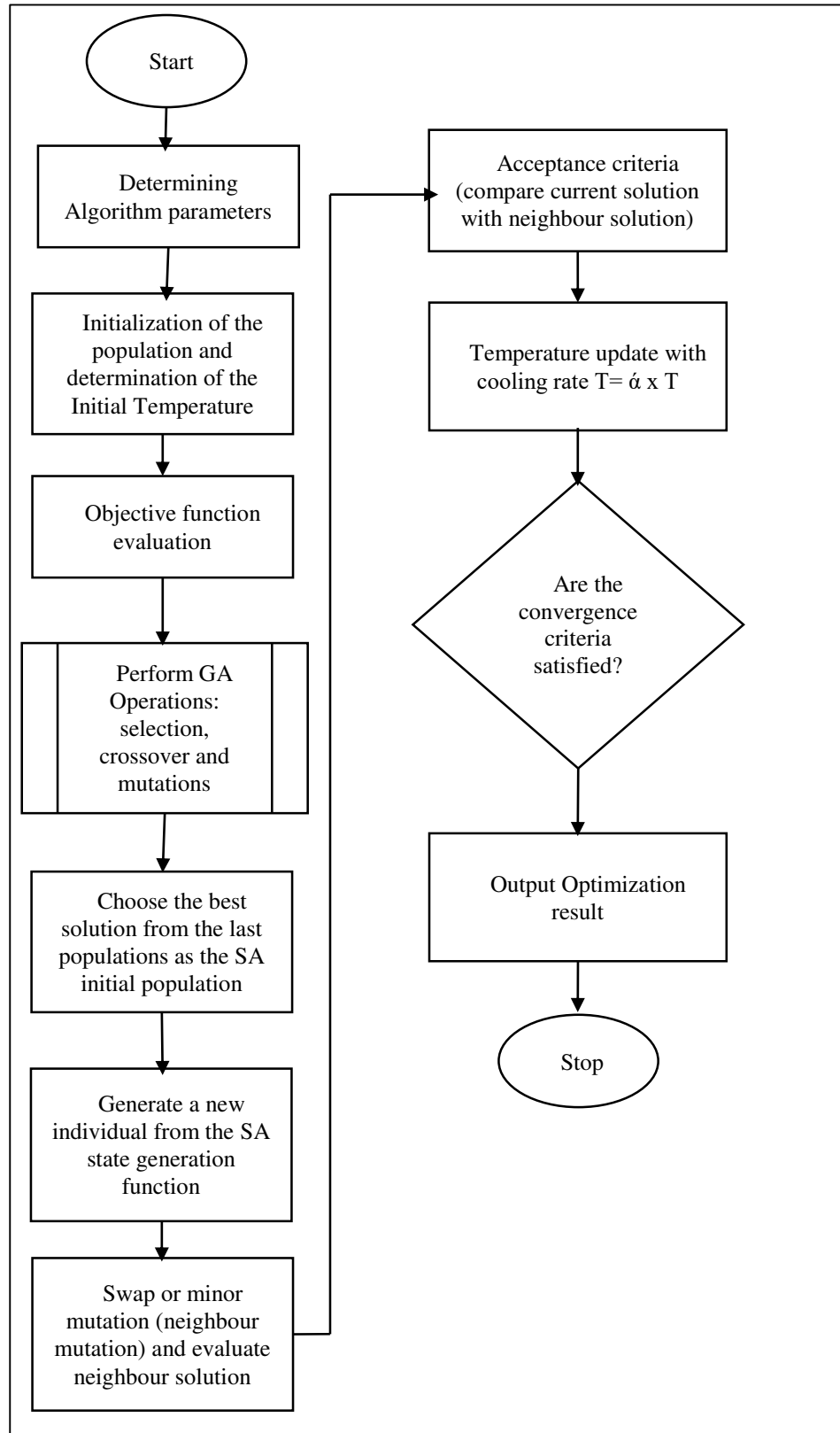


Figure 3. HGASA Process [28]

4. Stopping Criteria

Determining stopping criteria is an important step in the design of genetic algorithms. This criterion will determine when the algorithm should be stopped, either after achieving the desired result or when it meets the predetermined limits. In this study the discontinuation criterion is carried out if it has met the specified maximum number of iterations and If there is no significant change in the population match value over several generations, the algorithm can be stopped, indicating that there is no further progress. The determination of the parameters to be used is carried out by testing the number of populations, crossover probabilities, mutation probabilities, cooling factor and Initial temperature in the HGASA algorithm before conducting simulations in case studies.

IV. RESULT

1. Figures Overview of Distribution Cases in 2018 Natural Disasters in Sigi Regency

When a natural disaster occurs, Sigi Regency uses two locations as shelters for aid items before being distributed to natural disaster locations. The two

locations namely the Sigi Regency BPBD Office in Bora and the Sigi Regency BAPPEDA office in Mpanau. Based on this, the total cost and maximum travel time carried out during natural disasters are then calculated. The result show in **Table 1**.

<i>Number Shelter</i>	<i>Total Cost (IDR)</i>	<i>Maximum Travel Time (hr)</i>
2	409531721	27.575

Table 1. Results of Calculation of Total Cost and Maximum Travel Time

The results of the calculation were obtained with the provision that the distribution was carried out in accordance with the request from the disaster-affected areas or distribution points and deliveries using pickups, trucks and helicopters for several points that could not or were difficult to reach by land. Deliveries of aid items to distribution points are delivered according to the availability of aid items and one-way or adjacent lanes will be delivered at the same time, if aid items are available. The following relief routes in the event of a natural disaster by two shelters are shown in **Fig. 4** and **5** below.

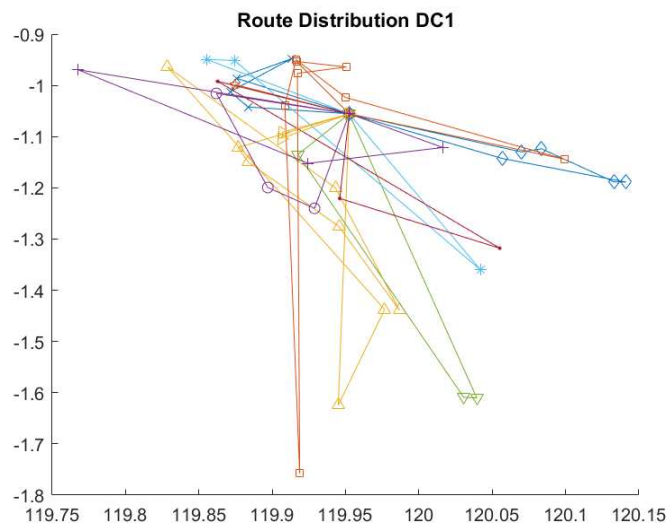


Figure 4. Distribution Route of Natural Disaster Assistance in Sigi Regency by shelter 1

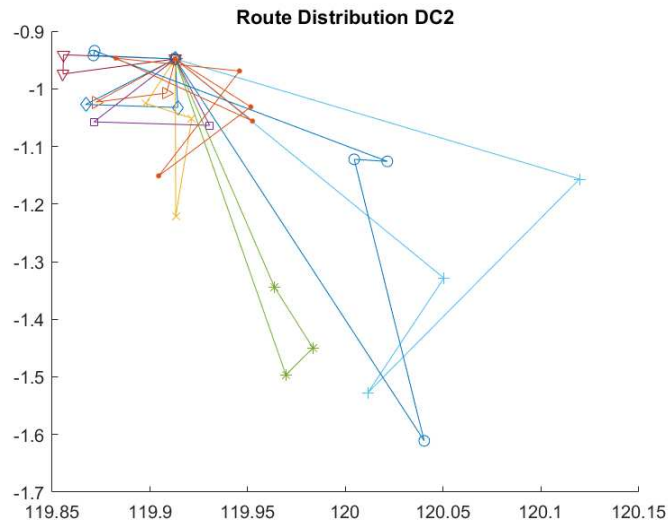


Figure 5. Distribution Route of Natural Disaster Assistance in Sigi Regency by shelter 2

2. Application Hybrid GASA Method

There are two scenarios for opening the number of open DCs, namely two DCs and three DCs. The simulation was carried out 56 times, of which 21 times for a combination of two DCs were opened and 35 times for a combination of three DCs were opened. The following are the results of the HGASA algorithm simulation shown in Fig. 6.

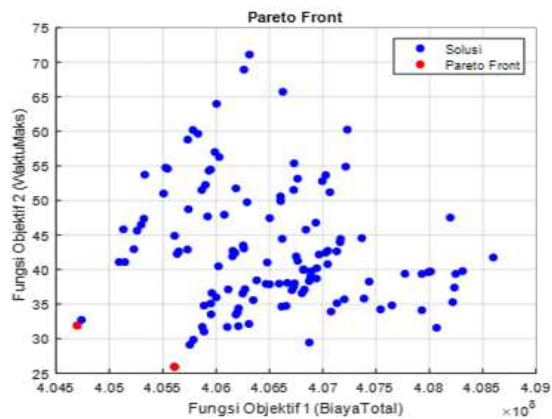


Figure 6. HGASA Results for 2 DC

Fig. 6 shows the distribution of solution results for the two distribution centres that were opened. The blue dots on the graph represent all the solutions generated by the HGASA algorithm, where each dot is a single solution that satisfies some combination of both objective functions. The red dot indicates the Pareto Front. The value range of objective function 1 ranges from Rp404,697,000 to Rp408,600,000 and the value range of the injective function 2 ranges from 25,921 hours to 71,124 hours. The solutions are scattered randomly on the chart, with some Pareto solutions prominent in the bottom left of the

chart. There are two pareto front solutions that are produced as the optimal solution, namely the combination of distribution centres 1,3 and 3,5.

Furthermore, the results of HGASA in the simulation for three DCs were opened, as shown in Fig. 7.

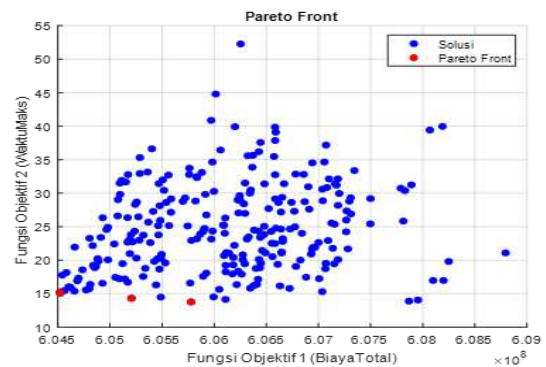


Figure 7. HGASA Results for 3 DC

Fig. 7 is a distribution of the solution results for the three distribution centers that were opened. The blue dots on the graph represent all the solutions generated by the HGASA algorithm, where each dot is a single solution that satisfies some combination of both objective functions. The red dot indicates the Pareto Front. The value range of objective function 1 ranges from IDR 604,519,000 to IDR 608,795,000 while the value range of objective function 2 ranges from 13,763 hours to 52.28 hours. The solution is spread out in a higher density around the middle value of Objective Function 1, with some Pareto solutions prominent in the bottom left of the graph. There are three pareto front solutions that are produced as optimal solutions, namely 1, 4 and 5, 3.4 and 5, 3.5 and 6.

From the results of the simulation, then the location

of the DC to be selected was determined. Decision making uses the Displaced Ideal Solution (DIS). The DIS method selects the best solution (W^*) from each existing solution (W_y), then the normalization process is carried out according to the equation that has been determined in the previous chapter. The normalization value can be negative if the solution being compared has a value greater than its ideal

value. However, when calculating the direct distance, use the absolute value of the normalized result to ensure the distance remains positive. The selection of the best decision is based on the smallest direct distance. The following are the results of the solution generated using the HGASA method shown in **Table 2**.

Table 2. HGASA optimization results

Number DC	DC Opened	Distribution Point	Route	Total Cost (IDR)	Max. travel time (hours)
2	3,5	D3: 1-2-3-5-7-8-10-11-12-13-14-17-20-21-22-24-25-27-29-33-34-35-36-38-39-40-41-42-44-45-48-49-53-54-56-57-59-60-61-62-63-66-67-70	Heli: 35-35-33 Truck: 56-59-1-54-60-36-21 Pickup: 53-12-14-67-62-41-20-7-24-38-48-39-49-10-22-57-13-17-61-5-25-3-42-63-29-66-8-70-44-40-2-27-11-45	405,609.300	25.921
		D5: 4-6-9-15-16-18-19-23-26-28-30-31-32-37-43-46-47-50-51-52-55-58-64-65-68-69-71-72	Heli: 30-72-52 Truck: 4-37-55-19-28-23-26-6-15-58-18-16-69-64-65-47-51-43-50-68-71-31-9-32-46 Pickup: -		
3	1,4,5	D1: 3-4-12-18-21-26-27-29-31-32-33-34-35-36-38-44-47-49-59-61-63-64-65-66-69-71	Heli: 34-35-33 Truck: 59-36-4-18-12-38-26-21-3-69-63-66 Pickup: 29-64-49-65-61-44-71-47-27-31-32	605,777,600	19.634
		D4: 1-2-7-8-11-14-15-16-19-23-24-25-28-39-40-41-42-45-46-53-54-55-57-60-70	Heli: - Truck: 54-60-53-1-14-41-7-24-19 Pickup: 55-39-46-57-28-15-42-16-8-23-25-70-40-2-11-45		
		D5: 5-6-9-10-13-17-20-22-30-37-43-48-50-51-52-56-58-62-67-68-72	Heli: 30-72-52 Truck: 56-67-62-17-20-37-10-6-5-58-51-43-50-68-48-22-13-9 Pickup: -		

The results of **Table 2**. show the opening of two DCs (DC3 and DC5), the total cost incurred is Rp 405,609,300 with a maximum distribution time of 25,921 hours. Although this cost is slightly lower compared to the two-DC scenario of NSGA II, the maximum distribution time is much higher, indicating a significant trade-off between cost and time efficiency. In DC3, heli routes are used to reach hard-to-reach distribution points, while trucks and pickups are optimized to reach areas with better

accessibility. In contrast, the DC5 demonstrates the effective use of heli and truck without the need to use pickups, indicating a simpler and more focused route.

In the scenario of three DCs (DC1, DC4, and DC5), the total cost incurred is IDR 605,777,600 with a maximum distribution time of 19,634 hours. This scenario combines the strategic use of helicopters, trucks, and pickups, where helicopters are used at specific distribution points in DC1 and

DC5 to optimize time. DC4, which does not use a heli, utilizes trucks and pickups to reach distribution points in more detail, demonstrating flexibility in addressing complex distribution challenges.

The use of vehicle routes shows significant adjustments to improve efficiency and reduce costs. The use of helicopters has remained consistent for certain areas that are inaccessible or difficult to reach by land, while trucks are used effectively to reach large areas with more focused routes. Pickups, which are used extensively in some DCs, allow for

distribution handling in denser areas or with more limited access.

3. Comparison of Overview of Distribution Cases in 2018 with the application of the HGASA Method

The results of the calculation of the overview of distribution case in Natural disaster 2018 in Sigi Regency and the application of HGASA is shown in **Table 3**.

Table 3. Result Overview of Distribution Cases in 2018 with the application of the HGASA Method

	<i>Number DC Opened</i>	<i>Total Cost (IDR)</i>	<i>Maximum time travel (hours)</i>
Distribution in Sigi (Natural Disaster 2018)	2	409,531,721	27.572
HGASA	2	405,609.300	25.921
	3	605,777,600	19.634

The calculation results presented in **Table 3** highlight significant differences between the humanitarian aid distribution in Sigi Regency in 2018 and the application of the HGASA method for varying numbers of distribution centres (DCs). In the 2018 humanitarian aid distribution in Sigi Regency, which did not utilize an optimization method, two shelters were opened, total cost reaches IDR409,531,721 and the maximum travel time is 27,572 hours. Meanwhile, using the HGASA method for two DCs that are opened, the total cost is IDR 405,609,300 and the maximum travel time is 25,921 hours, and for the opening of three DCs, the total cost is IDR 605,777,600 and the maximum travel time is 19,634 hours.

HGASA with 2 DCs offers cost savings over distribution in 2018. However, the addition of DC to 3 led to a significant increase in cost. HGASA, both

with 2 DC and 3 DC, improves travel time efficiency. The addition of DC to 3 provides a greater reduction in maximum travel time. If the main goal is to reduce the maximum travel time for a faster response, especially in emergency situations, opening 3 DCs with HGASA is more recommended. Hybrid Approach in HGASA Although HGASA uses a combination of genetic algorithms and simulated annealing, which in theory can provide an advantage in the exploration of solutions, some studies have shown that hybrid methods often have weaknesses in maintaining population diversity. Coello (2010) stated that hybrid algorithms can lose important solution diversity, making them more susceptible to convergence on local solutions. The following **Fig. 8** and **9** show route drawing for HGASA method Scenario 2 DC's. Then **Fig. 10** are route drawing for HGASA method scenario 3 DC's.

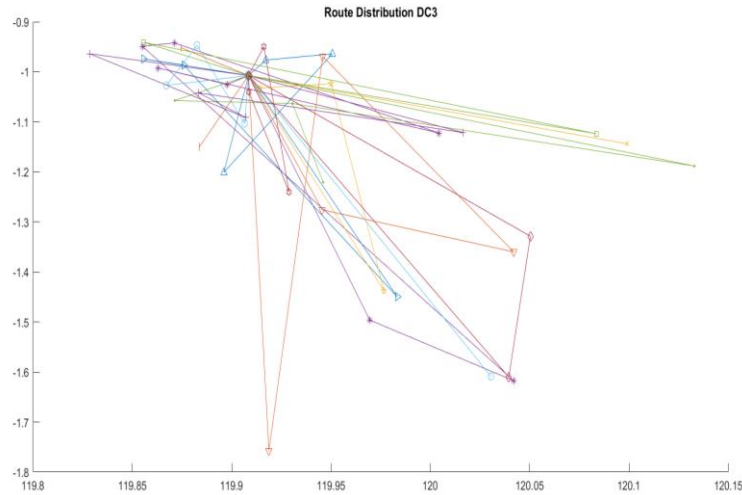


Figure 8. Distribution Route of HGASA Method Scenario 2 DC's by DC3

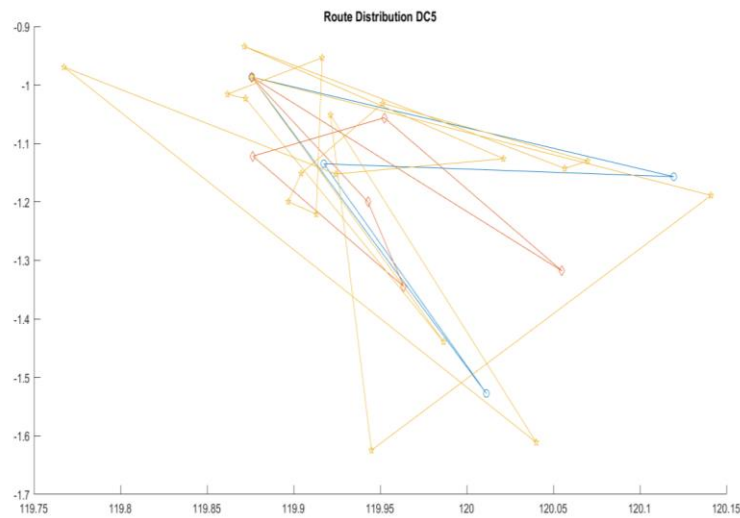


Figure 9. Distribution Route of HGASA Method Scenario 2 DC's by DC 5

Based on **Fig. 8** and **9** show that the distribution route is more centralized at two distribution centre points (DCs). Most routes have a more complex distribution pattern with some longer routes. The route network appears denser and overlaps each other, indicating that multiple receiving locations rely on a single DC, thus increasing the mileage. even some locations that are far from DC still must be served by one of the two DCs, thus increasing travel time. **Fig. 10** shows that the opening of the three DCs results in more spread routes, reducing the average length of the routes and creating a more efficient distribution. Route networks look more

segmented with better load sharing, reducing route overlap. locations that were far away from the previous two DCs can now be more easily reached by the third DC, reducing the average distance for distribution.

HGASA combines the global search capabilities of genetic algorithms (GA) and the local search capabilities of simulated annealing (SA). In this case, GA helps explore the solution space to find the optimal DC configuration (e.g., 2 DC or 3 DC), while SA fines out the solution by finding the best distribution route around the selected configuration.

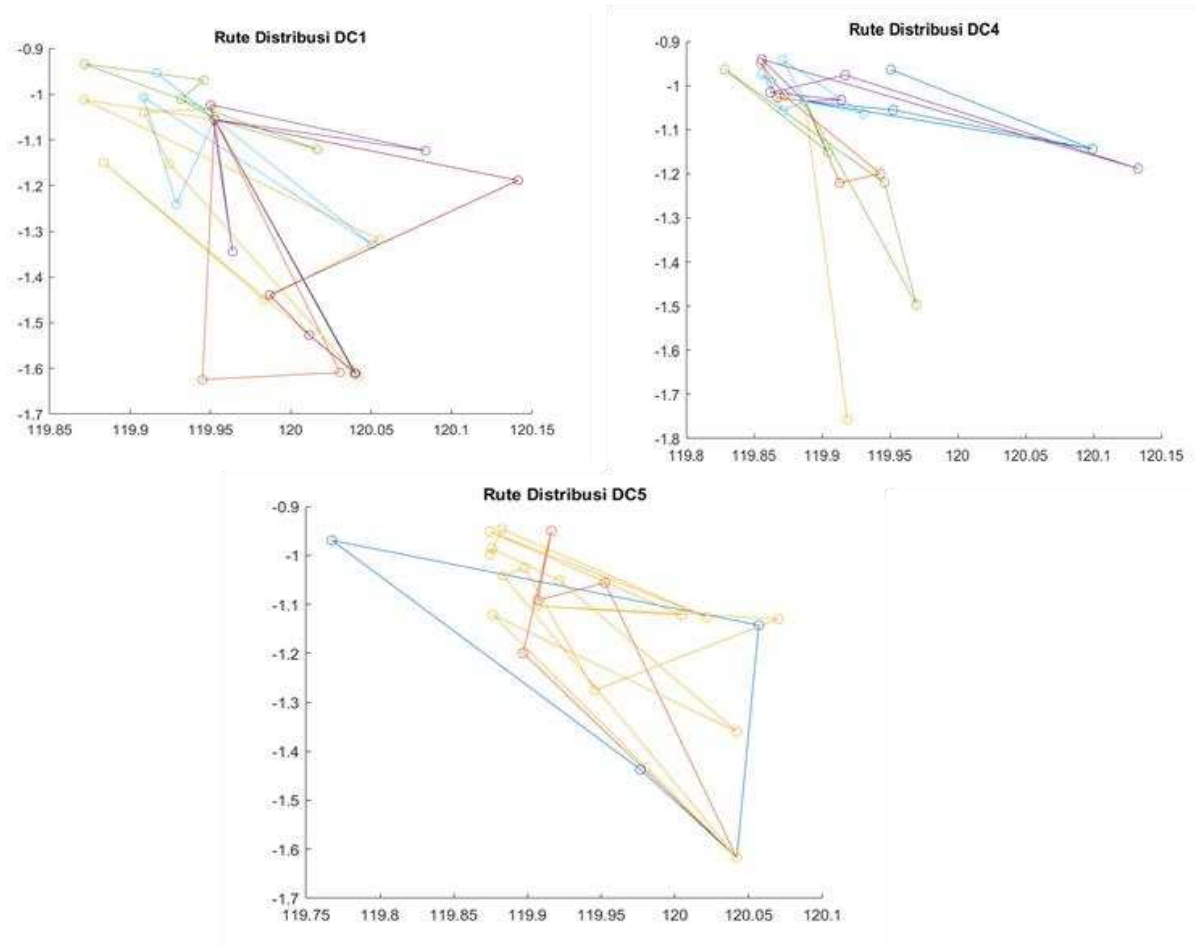


Figure 10. Distribution Route of HGASA Method Scenario 3 DC

V. CONCLUSION

Based on the results of research that has been carried out in the use of the HGASA method for the distribution of humanitarian aid using historical data of natural disasters in 2018 in Sigi district. The optimization of HGASA obtained two distribution centers can be carried out at the location of D3 and D5 or the location of the South Dolo Health Center in Balluase and the South Kulawi Sub-district Office in Tompi Bugis with a total cost of IDR 405,609,300 and a maximum service time of 25,921 hours while the construction of three distribution centers can be carried out at the location of DC1, DC4 and DC 5 or the location of the Sigi Bora BPBD Office, PUSDALOPS Maku Office and South Kulawi Sub-district Office in Tompi Bugis with a total cost of IDR 605,777,600 and a maximum service time of 19,634 hours.

The opening of three DCs by HGASA drastically reduces the maximum travel time, which can be a critical factor in emergency situations. If reducing travel time is a top priority (for example, for disaster emergency response), opening three DCs with HGASA is more effective. However, if cost control

is the primary focus, HGASA with two DCs is more optimal than the previous distribution.

Professional contributions from this publication namely (i) utilization of the HGASA method for disaster logistics management: this publication provides an innovative approach in optimizing the distribution of humanitarian aid by integrating historical data of disasters. This can help professionals in the field of logistics and disaster management in more effective and efficient strategic planning; (ii) reduced response time in emergency situations: this study provides a framework that government agencies and non-governmental organizations can apply to accelerate the delivery of aid, which is crucial in saving lives in emergency situations; and (iii) contribution to the academic literature: this research has become one of the important references in the application of the HGASA method for disaster logistics management, providing new insights for further research in this field.

Further research may focus on the development of more complex algorithms, testing in a variety of real-world conditions, and integration with modern technologies such as IoT and decision support

systems. This will ensure that the HGASA method not only provides an optimal theoretical solution but is also relevant for practical applications in a variety of disaster and logistics distribution scenarios.

AUTHOR CONTRIBUTIONS

S. Bombang: Conceptualization, Experiments, Theoretical analysis and writing.

B. P. Icthiarto: Supervision and Review.

H. H. Purba: Review and editing.

REFERENCES

- [1] UU, Undang-Undang Republik Indonesia Nomor 24 tahun 2007, 2007.
- [2] CRED, Disaster Year in Review 2023, no. 74, Apr. 2024.
- [3] BNPB, Rencana Nasional Penanggulangan Bencana 2020-2024, 2020.
- [4] BP3D, BPBD, and UNDP, Rencana Rehabilitasi dan Rekonstruksi: Paska bencana gempa, likuifaksi dan Bencana Lainnya, 2019.
- [5] D. Triatmaja, Dewanti, S. Irawan, Penentuan Loasi Warehouse dalam Mendukung Distribusi Kemanusiaan, *Seminar Nasional Energi dan Teknologi*, 2018.
- [6] X. Li, H. Yin, F. Yan, Routing optimization of the emergency supplies distribution vehicles using NSGA-II algorithm: a case study, *MATEC Web of Conferences*, vol. 325, p. 03002, 2020.
<https://doi.org/10.1051/mateconf/202032503002>
- [7] Milawaty, Isu Sosial Pascabencana Alam: Studi Kasus Penjarahan Di Kota Palu Melalui Pendekatan Pinheiro Principles Dan Psikologi Sosial, *Jurnal Administrasi Publik* vol 16, no 2, 2020.
<https://doi.org/10.52316/jap.v16i2.48>
- [8] C. Liu, G. Kou, Y. Peng, et al., Alsaadi, Location-routing problem for relief distribution in the early post-earthquake stage from the perspective of fairness,” *Sustainability*, vol. 11, no. 12, 2019.
<https://doi.org/10.3390/SU11123420>
- [9] S. Tri, W. Mara, A. M. S. Asih, and R. J. Kuo, Solving a Multi-Objective Location-Routing Problem with Minimum Cost and Total Time Balance, *International Journal of Information and Management Sciences*, vol. 30, pp. 249–269, 2019.
<https://doi.org/10.6186/IJIMS.201909>
- [10] J. Zhu, Non-linear Integer Programming Model and Algorithms for Connected p-facility Location Problem, *Journal of Systems Science and Information*, vol. 2, no. 5, pp. 451–460, Sep. 2017.
<https://doi.org/10.1515/jssi-2014-0451>
- [11] E. B. Tirkolaee, A. Goli, and G. W. Weber, A robust two-echelon periodic multi-commodity RFID-Based location routing problem to design petroleum logistics networks: A case study, *International Conference on Logistics*, 2020.
https://doi.org/10.1007/978-3-030-89743-7_1
- [12] F. R. Hamzani, S. Sitorus, S. Sitorus, et al., Optimization Model of Location Routing Problem for Disaster Relief Distribution, *Sinkron*, vol. 7, no. 3, pp. 2072–2079, Aug. 2022.
<https://doi.org/10.33395/sinkron.v7i3.11604>
- [13] A. P. Chobar, M. S. Ara, et al., A multi-objective location-routing problem model for multi-device relief logistics under uncertainty using meta-heuristic algorithm, *Journal of Applied Research on Industrial Engineering*, 2022.
<https://doi.org/10.22105/jarie.2021.299798.1365>
- [14] F. Zafari and D. Shishebori, Designing a Multi-Objective Three-Stage Location-Routing Model for Humanitarian Logistic Planning under Uncertainty, *Advances in Industrial Engineering*, vol. 2019, no. 4, pp. 149–167, 2019.
<https://doi.org/10.22059/jieng.2021.313355.1744>
- [15] H. Beiki, S. M. Seyedhosseini, V. R. Ghezavati, et al., A location-routing model for assessment of the injured people and relief distribution under uncertainty. *International Journal of Engineering, Transactions A: Basics*, vol. 33, no. 7, pp. 1274–1284, Jul. 2020.
<https://doi.org/10.5829/ije.2020.33.07a.14>
- [16] L. Shen, F. Tao, Y. Shi, and R. Qin, Optimization of location-routing problem in emergency logistics considering carbon

DISCLOSURE STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID

S. Bombang <http://orcid.org/0009-0004-6327-5201>

H. H. Purba <http://orcid.org/0000-0002-8166-6845>

- emissions, *Environmental research and Public Health*, vol. 16, 2019.
<https://doi.org/10.3390/ijerph16162982>
- [17] D. Veysmoradi, B. Vahdani, et al., Multi-objective open location-routing model for relief distribution networks with split delivery and multi-mode transportation under uncertainty, *Scientia Iranica*, vol. 25, no. 6E, pp. 3635–3653, 2018.
<https://doi.org/10.24200/sci.2017.4572>
- [18] Z. Liang, M. Liu, P. Zhong, et al., Hybrid Algorithm Based on Genetic Simulated Annealing Algorithm for Complex Multiproduct Scheduling Problem with Zero-Wait Constraint, *Math Probl Eng*, 2021.
<https://doi.org/10.1155/2021/9951995>
- [19] G. A. Fanshuri Alfarisy, A. Nugroho Sihananto, et al., Hybrid Genetic Algorithm and Simulated Annealing for Function Optimization, *Journal of Information Technology and Computer Science*, 2017.
<https://doi.org/10.25126/jitecs.20161215>
- [20] P. Xu, S. Sui, and Z. Du, Application of Hybrid Genetic Algorithm Based on Simulated Annealing in Function Optimization, *World Academy of Science, Engineering and Technology International Journal of Mathematical and Computational Science*, vol. 9, 2015.
- [21] X. Long, S. Wu, X. Wu, Y. Huang, and Z. Mu, A GA-SA hybrid planning algorithm combined with improved clustering for LEO observation satellite missions, *Algorithms*, vol. 12, no. 11, Nov. 2019.
<https://doi.org/10.3390/a12110231>
- [22] A. Sanagooy Aghdam, M. A. Afshar Kazemi, and A. Toloie Eshlaghy, A Hybrid GA-SA Multiobjective Optimization For RFID Network Planning Problem, *Journal of Applied Research on Industrial Engineering*, vol. 8, 2021.
<https://doi.org/10.22105/jarie.2021.295762.135710.2>
- [23] H. Wei, S. Li, H. Jiang, et al., Hybrid genetic simulated annealing algorithm for improved flow shop scheduling with makespan criterion, *Applied Sciences*, vol. 8, no. 12, Dec. 2018.
<https://doi.org/10.3390/app8122621>
- [24] M. Hajghani, M. A. Forghani, A. Heidari, et al., A two-echelon location routing problem considering sustainability and hybrid open and closed routes under uncertainty, *Heliyon*, vol. 9, no. 3, Mar. 2023.
<https://doi.org/10.1016/j.heliyon.2023.e14258>
- [25] T. Yan, F. Lu, S. Wang, L. Wang, and H. Bi, A hybrid metaheuristic algorithm for the multi-objective location-routing problem in the early post-disaster stage, *Journal of Industrial and Management Optimization*, vol. 19, no. 23, pp. 4663–4691, 2023.
<https://doi.org/10.3934/jimo.2022145>
- [26] A. Heidari, D. M. Imani, M. Khalilzadeh, et al., Green two-echelon closed and open location-routing problem: application of NSGA-II and MOGWO metaheuristic approaches, *Environ Dev Sustain*, vol. 25, no. 9, pp. 9163–9199, Sep. 2023.
<https://doi.org/10.1007/s10668-022-02429-w>
- [27] M. M. Munir, A. Pujianto, H. Aulia, et al., Optimisasi Algoritma Genetika dengan Particle Swarm Optimization (PSO) untuk Sistem Rekomendasi Diet Gizi bagi Penderita Diabetes, *Jurnal RESTIA*, vol. 1, no. 2, 2023.
<https://doi.org/10.30787/restia.v1i2.1289>
- [28] S. Goudarzi, W. H. Hassan, M. H. Anisim, et al., Comparison between hybridized algorithm of GA-SA and ABC, GA, DE and PSO for vertical-handover in heterogeneous wireless networks, *Sadhana - Academy Proceedings in Engineering Sciences*, vol. 41, no. 7, pp. 727–753, Jul. 2016.
<https://doi.org/10.1007/s12046-016-0509-4>



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license.

Research Article

Satellite-Based Validation of Contrail Prediction Models for Sustainable Aviation

Baddireddi Sree Chandana¹, Pelleti Nandieswar Reddy¹, Rithvika Alapati¹, Sai Aswath Reddy¹, Radha Doraisamy^{1, *}, Uma Sankari²

¹ Department of Computer Science and Engineering, Amrita School of Computing,
Amrita Vishwa Vidyapeetham Bengaluru Campus Kasavanahalli, Carmelaram P.O.
Bengaluru - 560 035 Karnataka, India

² Vikram Sarabhai Space Centre, GV6Q+XCG, Veli - Perumathura Rd,
Kochuveli, Thiruvananthapuram, Kerala 695022, India

*e-mail: d_radha@blr.amrita.edu

Submitted: 21/12/2024 Revised: 17/02/2025 Accepted: 20/02/2025 Published online: 26/02/2025

Abstract: The aviation industry significantly contributes to global warming through the formation of contrails, which trap heat in the atmosphere and exacerbate climate change. To mitigate this effect, sophisticated models have been developed to predict contrail formation and its associated warming effects, but these require empirical validation for accuracy. This project leverages satellite imagery to validate contrail prediction models, enabling effective contrail avoidance strategies for airlines. U-Net variants, a convolutional neural network architecture, is utilized for image segmentation to identify contrails in satellite imagery. By optimizing the threshold for the softmax layer, contrail detection accuracy, and validating model predictions with real-world data had been enhanced. This enables pilots to minimize contrail formation during flights, aiming to reduce the aviation industry's environmental impact. The research offers a scalable and cost-effective solution for enhancing aviation sustainability and aligns with global efforts to combat climate change.

Keywords: *Contrails; U-Net; Image Segmentation; Satellite Imagery; Climate Change; Contrail Prediction Models*

I. INTRODUCTION

The aviation industry has long been recognized as a significant contributor to global warming, primarily due to the formation of contrails – those line-shaped clouds of ice crystals emitted from aircraft engine exhaust. These contrails have been identified as a key player in climate change dynamics, exacerbating the environmental impact of air travel by trapping heat in the atmosphere. As concerns over climate change intensify, it becomes imperative for the aviation sector to address its environmental footprint.

Contrails, which are essentially artificial clouds, are formed whenever hot, moist exhaust from aircraft engines mixes with cold air at high altitudes. They not only contribute to the visual pollution of the skies but also play a crucial role in altering the Earth's radiation balance, leading to increased warming of the atmosphere. This phenomenon is particularly concerning given the exponential growth of air travel worldwide, with projections

indicating continued expansion in the coming decades. Recognizing the urgency of mitigating the aviation industry's impact on climate change, researchers have devoted significant efforts to understanding and predicting contrail formation. Sophisticated models have been developed to simulate the complex interactions between aircraft emissions, atmospheric conditions, and contrail formation processes. These models serve as valuable tools for assessing the environmental impact of air travel and devising strategies to minimize it. However, despite advancements in modelling techniques, there remains a critical need for empirical validation to enhance the accuracy and reliability of contrail prediction models. While laboratory experiments and field measurements provide valuable insights, they are often limited in scale and scope. Moreover, conducting extensive observations in the atmosphere poses logistical challenges and may not capture the full complexity of contrail formation dynamics. To address this gap, this project proposes leveraging satellite imagery as

a complementary approach to validate contrail prediction models. Satellites offer a unique vantage point from which to observe contrail formation on a global scale, providing comprehensive coverage of flight paths and atmospheric conditions. By correlating model predictions with real-world observations obtained from satellite imagery, researchers can validate and refine their understanding of contrail formation processes.

The integration of satellite data into contrail prediction models holds promise for enhancing the accuracy of forecasts and improving the effectiveness of contrail avoidance strategies. Armed with validated models, airlines can empower pilots to make informed decisions during flight planning and operations, taking into account factors such as optimal altitude, route selection, and engine settings to minimize contrail formation. By doing so, the aviation industry can mitigate its contribution to climate change while ensuring the sustainability of air travel. In alignment with global efforts to combat climate change, this research aims to provide a scalable and cost-effective solution for reducing the environmental impact of the aviation sector. By leveraging satellite imagery to validate contrail prediction models, this project seeks to enable more informed decision-making and promote sustainable practices within the aviation industry. Through collaborative efforts between researchers, policymakers, and industry stakeholders, it is aimed to work towards a future where air travel is both efficient and environmentally responsible.

II. RELATED WORK

Contrails, or condensation trails, form behind aircraft and significantly contribute to aviation-induced climate change. The mitigation of their impact through contrail avoidance strategies is considered a cost-efficient method to reduce aviation's climate footprint. The introduction of the Open Contrails dataset aims to facilitate the development and evaluation of contrail detection models. This dataset, comprising manually labelled imagery from the GOES-16 Advanced Baseline Imager (ABI), is designed to train models capable of identifying contrails with high accuracy [1]. A proposed contrail detection model incorporates temporal context, enhancing detection accuracy by integrating temporal information into CNN(convolution neural network)-based models. This approach shows promise for better understanding contrail dynamics. The study also suggests leveraging self-supervised and semi-supervised learning techniques to further improve model performance. Expanding the research to include data from geostationary satellites like Himawari-8 and Meteosat-11 could extend coverage, especially over Europe and the Asia-Pacific region, highlighting the importance of

satellite data and CNN models in addressing aviation-induced climate change [1].

A comprehensive study focuses on the detection, tracking, and matching of linear contrails using geostationary satellite infrared images, weather data, and air traffic data. The primary objective is to create a dataset that captures the complete lifecycle of contrails and identifies the aircraft responsible for their formation. This innovative methodology simultaneously addresses tracking and identification challenges, providing a holistic and integrated approach to contrail analysis. The integration of satellite data with air traffic information offers a robust platform for investigating the environmental impact of contrails, enabling real-time monitoring and improved air traffic management [2]. Machine learning techniques have been applied to identify contrails in images captured by the United States Department of Energy's Atmospheric Radiation Management (ARM) user facility. A deep convolutional neural network trained on 1600 photos from the Total Sky Imager (TSI) achieved high accuracy rates. Another study used a CNN to distinguish contrail cirrus clouds from regular cirrus clouds, employing Python packages for implementation and achieving notable performance through binary cross-entropy loss and adaptive momentum optimization [3].

A novel method for contrail identification in satellite images employs semantic segmentation, utilizing the UPerNet architecture with ConvNeXt configurations. This model effectively handles class imbalances and uses the AdamW optimizer for fine-tuning, achieving outstanding performance. This approach underscores the potential for improved contrail identification in satellite imagery through advanced segmentation techniques [4]. Researchers developed a CNN specifically tailored for contrail detection in satellite imagery, yielding promising results with a probability of detection at 0.51, a false alarm ratio of 0.46, and an F1 score of 0.52. The CNN's impressive performance, evidenced by an AUC-PR of 73.9, highlights its potential for large-scale contrail monitoring and a better understanding of their climate impacts [5]. The need for accurate and automated contrail detection algorithms has led to the development of deep segmentation models for contrail detection in Landsat-8 imagery. UNet with Xception 71 as the encoder backbone performed best, achieving an IoU of 0.4395. Despite challenges, this study represents significant progress in using advanced segmentation methods for contrail detection [6].

To address label bias in contrail identification, a probabilistic deep learning approach using P-UNet is proposed. This method shows resilience to label biases and improves recall, suggesting robustness and generalizability across diverse satellite image datasets. Future research could enhance precision

while maintaining high recall, potentially incorporating additional contextual information or refining labelling methodologies [7]. A novel approach based on few-shot transfer learning, using pre-trained segmentation models and the SR Loss function, significantly improves contrail detection performance. This method overcomes challenges posed by limited labelled datasets and varied image conditions, offering a robust solution for contrail detection in remote sensing imagery [8]. Research on sky imaging for solar radiation estimation emphasizes cloud segmentation as a crucial step. A high-resolution cloud segmentation dataset created using sky images can facilitate future research in meteorology, weather forecasting, and solar energy forecasting. This dataset, consisting of 825 manually labelled sky photographs, enhances segmentation accuracy and supports various studies in related fields [9]. Contrail avoidance strategies require reliable models to be effective. This study compares two models, CoCiP (contrail cirrus prediction) and APCEMM (Aircraft Plume Chemistry, Emissions, and Microphysics Model), under various conditions to evaluate their accuracy and sensitivity. The findings highlight the need for more validation data and simple models that meet the minimum accuracy required for contrail prediction and avoidance, crucial for reducing aviation's environmental impact [10].

Recent studies have explored complementary approaches to contrail detection and analysis. For instance, radiosondes, widely used in atmospheric research, provide critical data for validating satellite-based contrail models. Improved radiosonde technology enhances atmospheric measurements, indirectly supporting contrail detection efforts [11]. Combining radiosonde-based atmospheric measurements with machine learning models can further refine contrail detection algorithms [12]. Additionally, safer radiosonde deployments reduce risks associated with atmospheric data collection, ensuring consistent and reliable inputs for contrail analysis [13]. These advancements underscore the importance of integrating multiple data sources to improve the robustness of contrail detection systems.

Another area of interest involves aerospace sustainability efforts. Recent research highlights the role of machine learning in optimizing flight paths to minimize contrail formation [14]. Such studies emphasize the need for interdisciplinary approaches that combine aerospace engineering, atmospheric science, and machine learning to address aviation-induced climate change. Furthermore, innovations in high-resolution imaging and sensor technologies have enabled more precise contrail detection, particularly in challenging environments [15]. These developments align with the broader goal of creating sustainable aviation practices.

In summary, the reviewed studies highlight significant advancements in contrail detection and mitigation strategies using satellite imagery and machine learning techniques. These approaches, from developing comprehensive datasets to leveraging advanced neural network architectures, contribute to reducing the aviation industry's environmental impact. Future research should focus on enhancing model accuracy, expanding datasets, and integrating additional contextual information to improve contrail detection and avoidance strategies. By drawing inspiration from related work and incorporating insights from complementary fields such as atmospheric science and aerospace engineering, higher accuracies can be achieved, and the robustness of contrail detection models can be enhanced.

III. SYSTEM DESIGN

A. High-Level Architecture

The contrail detection system is designed to process raw imagery data, which undergoes preprocessing steps such as normalization, noise reduction, and format conversion to ensure compatibility with the detection model. At the core of the system lies a U-net CNN and its variants, known for their effectiveness in image segmentation tasks. These models are trained on preprocessed imagery along with corresponding ground truth data, specifying the locations of contrails.

To address specific challenges and enhance performance, the system incorporates several variants of the U-net. These include the Attention UNet, Residual UNet, Attention Residual UNet, Attention Residual UNet with ELU(Exponential linear unit) Activation, and Attention Residual UNet with PreLU(parametric rectified linear unit) Activation. Each variant offers unique advantages such as attention mechanisms, residual connections, and advanced activation functions.

During the training phase, multiple iterations occur, with each U-Net variant trained on preprocessed data and ground truth labels. Evaluation metrics such as the dice score and pixel-wise accuracy are monitored to assess segmentation effectiveness. Visualizing model predictions aids in understanding performance and identifying areas for improvement. Following training, softmax thresholding is applied in post-processing to generate clear segmentation boundaries for contrail predictions, ensuring precise and accurate results. Trained models undergo testing on unseen imagery to validate performance across diverse datasets.

This comprehensive system leverages U-Net variants for automated contrail detection, integrating attention mechanisms, residual connections, and advanced activation functions to enhance model performance. Visualization, softmax thresholding,

rigorous training, and testing contribute to the system's robustness in detecting contrails in various imagery datasets. **Figure 1** illustrates the high-level architecture of the contrail detection system.

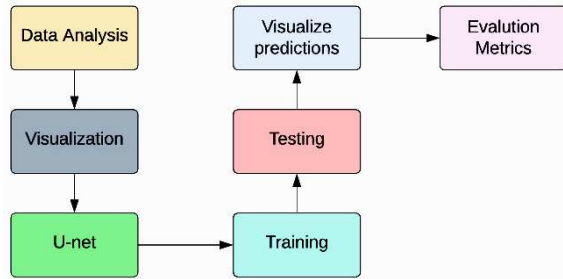


Figure 1. High-Level architecture

B. Low-Level Architecture

The low-level architecture of the contrail detection model encompasses a series of intricately designed components, each contributing to the model's efficacy and robustness. It begins with the data loading process, where a dataset comprising images and corresponding labels (masks) is imported into the system. Once the data is loaded, the next step involves preprocessing, which includes normalization and resizing of the images to ensure consistent input across all UNet variants. After preprocessing, the data is partitioned into training and validation sets. Having a separate validation set allows for the evaluation of the model's accuracy and facilitates informed decisions regarding its design and training process. This step is crucial to prevent overfitting to the training data, ensuring that the model generalizes well to unseen data.

The model architecture is then devised, incorporating the base UNet and various UNet variants tailored to enhance performance and address specific challenges. The Attention UNet integrates attention mechanisms within the skip connections to focus on relevant features while suppressing noise. This design improves information flow between the down-sampling and up-sampling paths, enabling the model to prioritize important areas.

Another variant, the Residual UNet, employs residual blocks in both the contracting and expansive paths. These residual connections enable the architecture to delve deeper, mitigating issues related to gradient flow and enhancing training stability. The Attention Residual UNet combines attention mechanisms with residual connections, resulting in a robust architecture capable of capturing complex structures while maintaining focus on pertinent details.

Additionally, the Attention Residual UNet with ELU Activation incorporates ELU activation function to expedite convergence and enhance training stability. ELU's positive output retention aids in preventing dead neurons, leading to smoother and faster training. Similarly, the Attention Residual

UNet with PReLU Activation utilizes PReLU to provide trainable negative slopes for more flexible activations, offering adaptability in complex image segmentation tasks. Each architecture variant undergoes training using the training dataset, with regular evaluations on the validation set to monitor progress and make necessary adjustments. Post-processing techniques like softmax thresholding are applied to generate clear segmentation boundaries for contrail predictions. Following training, the models are assessed using a separate test dataset to evaluate their generalization to unseen data. Ultimately, the architecture's effectiveness is determined based on evaluation metrics, with a comparative analysis conducted to identify the most suitable variant. This meticulous design process ensures that the contrail detection model achieves optimal performance and reliability in real-world applications. **Figure 2** illustrates the low-level architecture of the contrail detection system.

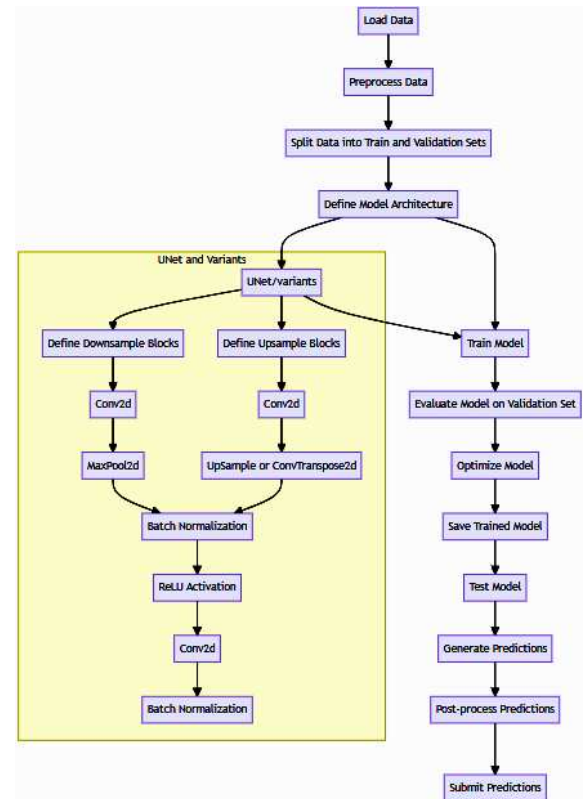


Figure 2. Low-Level architecture

IV. SYSTEM IMPLEMENTATION

The U-Net architecture, conceptualized by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in 2015, represents a pivotal milestone in the domain of biomedical image segmentation within the ambit of deep learning. It was specifically devised to surmount the challenges posed by sparse annotated data and the imperative to retain intricate spatial details in segmentation tasks, where conventional CNN architectures often faltered in preserving spatial coherence during the down-sampling process.

U-Net's innovative design circumvents these constraints through a symmetric encoder-decoder architecture endowed with skip connections, thus ensuring the seamless integration of both low-level and high-level features across varying spatial resolutions. The architecture is composed of three fundamental constituents: skip connections, the contracting path (encoder), and the expansive path (decoder).

The contracting path, constituting convolutional and pooling layers, progressively diminishes the spatial dimensions of the input image while concurrently augmenting the number of feature channels. Each convolutional block within the contracting path typically encompasses multiple convolutional layers, complemented by rectified linear unit (ReLU) activations and batch normalization, thereby facilitating feature extraction across hierarchical levels of abstraction.

Conversely, the expansive path leverages upsampling layers to restore the spatial dimensions of the feature maps while concomitantly reducing the number of channels. This restoration process is facilitated by transposed convolutional layers, colloquially referred to as deconvolution layers, which are intricately concatenated with feature maps originating from the contracting path. Such concatenation enables the faithful reconstruction of high-resolution feature maps while concurrently preserving crucial spatial contextual cues.

The quintessential feature of the U-Net architecture lies in its skip connections, which forge direct connections between analogous spatial resolutions within the encoder and decoder paths. By virtue of this architectural peculiarity, U-Net facilitates the fusion of localized spatial details gleaned from the contracting path with the holistic contextual information extracted by the expansive path, thereby endowing the model with the requisite acumen for precise and contextually informed segmentation.

In the realm of spatial dimension augmentation techniques, two prevailing methodologies predominate: up-sampling and convolution transpose. The former entails employing interpolation techniques such as nearest neighbor or

bilinear interpolation, thereby obviating the need for additional learnable parameters. Conversely, convolution transpose entails executing an inverse convolution operation, necessitating the learning of an additional set of trainable parameters.

Up-sampling garners favor within the domain of image segmentation tasks by virtue of its inherent simplicity, computational efficiency, and efficacy in artifact mitigation. Unlike convolution transpose, up-sampling methods do not engender the pernicious phenomenon of checkerboard artifacts, rendering them especially well-suited for tasks predicated upon accurate localization and delineation of object boundaries.

The salient advantages underpinning the efficacy of the U-Net architecture encompass its innate capability to preserve spatial information across the hierarchical depth of the network, its inherent adaptability to an eclectic array of segmentation tasks spanning diverse domains, and its modular design ethos, which endows it with the flexibility requisite for facile extension and customization in accordance with the exigencies of specific segmentation tasks and input modalities. These distinctive attributes collectively underscore the indelible imprint of U-Net as a preeminent architecture for image segmentation endeavors across a panoply of domains and applications.

Here are the key points explaining each layer of the U-Net architecture shown in **Figure 3**.

1. Input:

- Input images with dimensions 256x256 and 24 channels.

2. DoubleConv (inc):

- Applies two 3x3 convolutions with ReLU activation and batch normalization.

- Input: 256x256x24

- Output: 256x256x64

3. Down1:

- Downsampling via max-pooling followed by a DoubleConv block.



Figure 3. UNet architecture

- Input: 256x256x64
- Output: 128x128x128
- 4. Down2:
 - Down sampling via max-pooling followed by a DoubleConv block.
 - Input: 128x128x128
 - Output: 64x64x256
- 5. Down3:
 - Down sampling via max-pooling followed by a DoubleConv block.
 - Input: 64x64x256
 - Output: 32x32x512
- 6. Down4:
 - Down sampling via max-pooling followed by a DoubleConv block.
 - Input: 32x32x512
 - Output: 16x16x512
- 7. Up1:
 - Up sampling followed by concatenation with corresponding feature maps from Down3.
 - Input: 16x16x512 (from Down4), 32x32x512 (from Down3)
 - Output: 32x32x256
- 8. Up2:
 - Up sampling followed by concatenation with corresponding feature maps from Down2.
 - Input: 32x32x256 (from Up1), 64x64x256 (from Down2)
 - Output: 64x64x128
- 9. Up3:
 - Up sampling followed by concatenation with corresponding feature maps from Down1.
 - Input: 64x64x128 (from Up2), 128x128x128 (from Down1)
 - Output: 128x128x64
- 10. Up4:
 - Upsampling followed by concatenation with corresponding feature maps from inc.
 - Input: 128x128x64 (from Up3), 256x256x64 (from inc)
 - Output: 256x256x64
- 11. Output:

- Final convolution to map to the desired number of output channels (1 for binary segmentation).

- Input: 256x256x64

- Output: 256x256x1

This breakdown highlights the operations performed at each layer, including down sampling, up sampling, concatenation, and convolution, along with the changes in input and output dimensions. The tabular form of the layers is shown in **Table 1**.

Table 1. UNet layers architecture

Layer	Input Size	Output Size	Input Channels	Output Channels
Input	256x256x24	-	24	-
Inc	256x256	256x256	24	64
Down1	256x256	128x128	64	128
Down2	128x128	64x64	128	256
Down3	64x64	32x32	256	512
Down4	32x32	16x16	512	512
Up1	16x16	32x32	512	256
Up2	32x32	64x64	256	128
Up3	64x64	128x128	128	64
Up4	128x128	256x256	64	64
Output	256x256	256x256	64	1

The U-Net architecture presented above encompasses several key parameters and a dedicated trainer class is essential for its effective training. At the core of the architecture lie parameters that define its behavior and performance during the training process. These parameters include the configuration of the neural network itself, defined within the U-Net class. The architecture of the U-Net comprises various convolutional layers, down-sampling blocks, up-sampling blocks, and skip connections, all of which play crucial roles in capturing hierarchical features and preserving spatial information. Additionally, hyper parameters such as learning rate, batch size, and loss function are vital in guiding the optimization process and determining the model's convergence and performance.

Complementing the architecture, the custom trainer class orchestrates the training process by managing the flow of data, optimization, and evaluation. The class encapsulates essential functionalities such as forward pass computation, loss calculation, gradient computation, and parameter updates. It interfaces with the provided optimizer, loss function, and learning rate scheduler to optimize the U-Net model's parameters iteratively. Throughout the training process, the class keeps track of various metrics, including batch losses, epoch losses, learning rates, and validation losses, providing insights into the model's performance and convergence.

Furthermore, the class enables seamless integration with PyTorch's DataLoader module, facilitating efficient data loading and batching for

both training and evaluation. By iterating over the training and validation datasets, the trainer updates the model's parameters iteratively, adjusting the learning rate dynamically based on the specified schedule. Additionally, the trainer periodically evaluates the model's performance on the validation dataset, allowing for model checkpointing and monitoring of training progress.

In essence, the U-Net architecture and its associated trainer class form a cohesive framework for training and evaluating semantic segmentation models. Together, they leverage the power of deep learning to tackle complex image analysis tasks, offering a robust and adaptable solution for a wide range of applications, from medical imaging to remote sensing and beyond.

The Attention U-Net architecture builds upon the traditional U-Net model as shown in **Figure 4**, introducing attention mechanisms to enhance the effectiveness of skip connections. The core idea is to improve segmentation by emphasizing critical features during the merging of information between the encoder and decoder paths.

The base structure of U-Net includes two main pathways: an encoder for down-sampling and a decoder for up-sampling. To maintain high-resolution details, U-Net employs skip connections, where features from the encoder are concatenated with those of the decoder.

In the Attention U-Net, these skip connections are enhanced with attention mechanisms. The attention gates are introduced to refine the merging of encoder and decoder features by allowing the model to focus on specific regions that are most pertinent to the current work. This focus is achieved by learning a weighting function that assigns different levels of importance to various parts of the feature maps. The

attention gates take as input a combination of encoded features and up-sampled decoded features, creating an "attention signal." This signal is processed to produce attention coefficients, typically using linear transformations followed by non-linear activation functions like Sigmoid, resulting in a spatial map that indicates the significance of each region as shown in **Figure 5**.

These attention coefficients are then used to scale the features from the encoder before they're concatenated with the decoder's output in the skip connections. By doing this, the Attention U-Net can selectively highlight important features while suppressing less relevant or noisy data. This mechanism contributes to a more efficient and focused merging process, improving the model's ability to capture intricate structures and complex relationships within the data.

The benefits of this approach are substantial. Additionally, the attention mechanisms help reduce noise and unnecessary information, leading to improved generalization and robustness. Overall, this selective attention strategy provides the Attention U-Net with an edge in terms of performance and accuracy, especially in scenarios where detailed structures need to be accurately segmented from complex backgrounds.

The Residual U-Net architecture extends the standard U-Net by incorporating residual connections, a concept popularized by Res-Net. In traditional CNNs, deep architectures can lead to issues such as vanishing gradients and difficulty in training. Residual connections address these problems by adding a "shortcut" or direct path that skips one or more layers, allowing the gradient to flow more easily during backpropagation as shown in **Figure 6**. This can enable deeper networks with improved training dynamics.

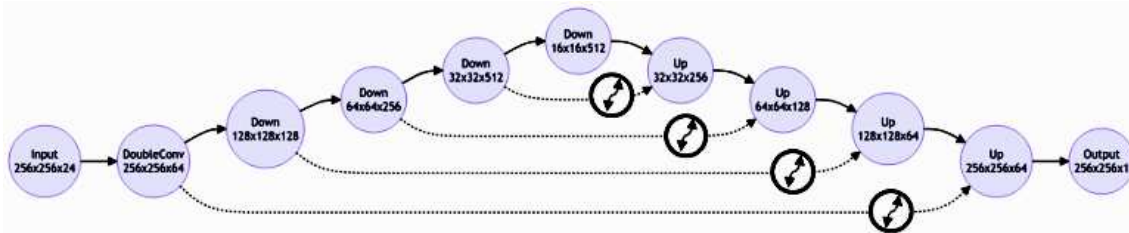


Figure 4. Attention Unet architecture

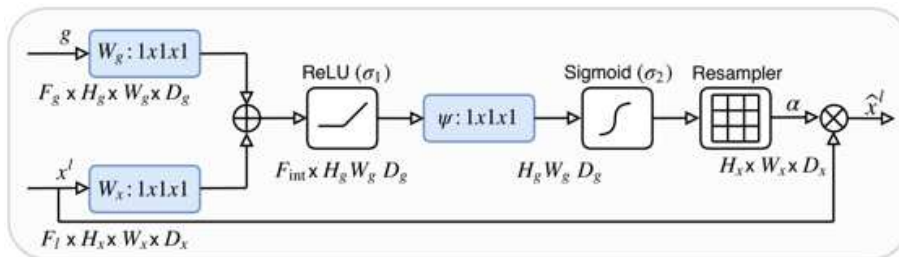


Figure 5. Attention mechanism

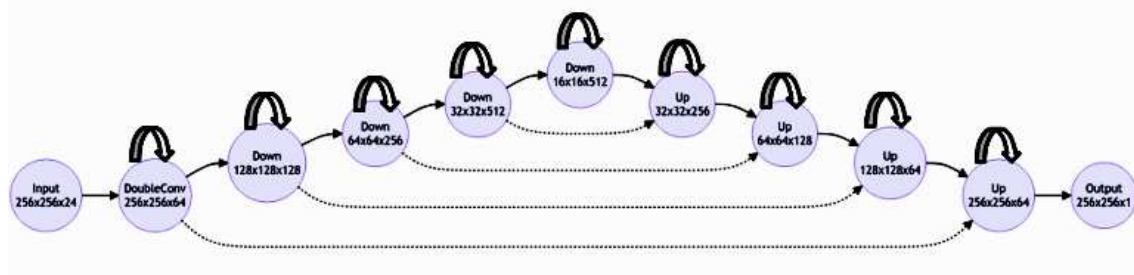


Figure 6. Residual UNet architecture

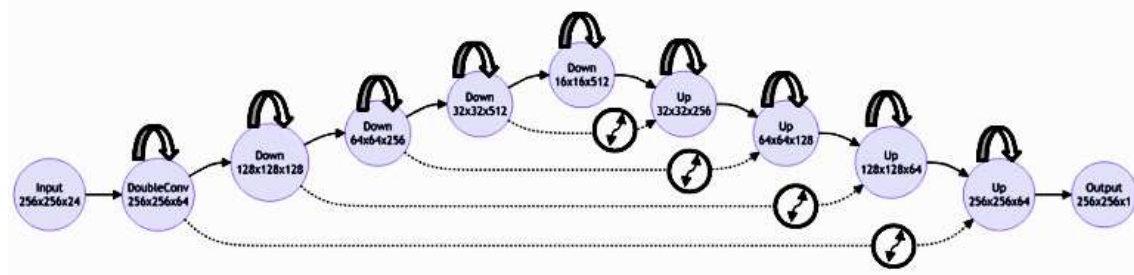


Figure 7. Attention Residual UNet architecture

In Residual U-Net, each block in the encoder and decoder contains residual connections. This means that rather than simply stacking convolutional layers, the input to each block is added to the output, creating a direct path for information flow. The encoder in Residual U-Net compresses the input through a series of convolutional blocks followed by down-sampling, usually with max-pooling. The decoder mirrors this structure but in reverse, using up-sampling techniques like transposed convolutions. The residual connections allow the model to focus on learning the residual changes rather than complete transformations, thereby promoting training stability and ease of learning. Skip connections are a hallmark of U-Net, allowing high-resolution features from the encoder to be directly linked with the corresponding stages of the decoder. In Residual U-Net, these skip connections can also contain residual connections, ensuring consistent information flow and reducing degradation in performance as the network gets deeper.

The Attention and Residual U-Net architecture shown in **Figure 7** combines the benefits of residual connections with those of attention mechanisms. The attention mechanisms, typically implemented through attention gates, let the model concentrate on the most relevant features in the skip connections, enhancing significant parts of the feature maps while downplaying less important areas. In this architecture, the encoder and decoder use residual connections, providing stability and deeper networks. The attention gates are integrated into the skip connections, where they take as input both the encoded features and the corresponding up-sampled features from the decoder. The attention gate then

produces a spatial attention map, indicating where the model should focus its attention. This map is used to scale the features from the encoder, allowing the model to selectively emphasize the most relevant information before concatenating with the decoder features.

This combination creates a powerful architecture for segmentation tasks. The residual connections ensure that the model can be trained efficiently without significant loss in performance as the network depth increases, while the attention mechanisms help the model focus on the most important aspects of the data.

The Attention and Residual U-Net with the ELU activation function enhances the architecture's stability and gradient flow. ELU is an activation function that introduces a smoother transition for negative inputs, allowing for small negative outputs. This characteristic can help reduce the vanishing gradient problem and improve training dynamics, especially in deep architectures like U-Net.

$$ELU(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha(e^x - 1), & \text{otherwise} \end{cases}$$

The ELU is an activation function used in neural networks to improve learning.

ELU stands out because it is smooth and differentiable everywhere, aiding in smooth gradient descent optimization. Unlike ReLU, which only outputs positive values, ELU allows negative values for negative inputs, helping balance the activations and reduce bias shifts. This feature, along with maintaining a small gradient for negative inputs, helps mitigate the vanishing gradient problem, making ELU particularly beneficial for deep

networks. Although ELU requires slightly more computational power due to the exponential calculation and necessitates tuning the hyperparameter α , it leads to faster and more stable learning, making it a valuable tool for training deep neural networks.

In this variant, the ELU activation function replaces the traditional ReLU in the convolutional layers. The encoder and decoder blocks, as well as the residual connections, use ELU, which leads to smoother gradients and improved stability during training. Combined with the attention mechanisms, this architecture can better focus on important features and maintain consistent gradient flow. This is particularly beneficial in scenarios requiring deep networks and high segmentation accuracy.

The Attention and Residual U-Net with PReLU activation function adds flexibility and adaptability to the architecture. An activation function called PReLU adds a learnable parameter to regulate the slope of the function's negative portion. Because of its adaptability, the model can change the activation function while training in response to the data.

The PReLU is an activation function used in neural networks, introduced to address some limitations of the standard ReLU. It enhances the learning capability by allowing a small, learnable slope for negative inputs, which helps to prevent the "dying ReLU" problem, where neurons can become inactive and stop learning. The PReLU activation function is defined as:

$$PReLU(x) = \begin{cases} x, & \text{if } x > 0 \\ ax, & \text{if } x \leq 0 \end{cases}$$

PReLU is particularly useful in deep learning architectures where flexibility in activation functions can help the network adapt better to the data. It is commonly used in CNNs and other deep networks where mitigating the risk of dead neurons and improving gradient flow are crucial.

In this variant, the PReLU activation function is used in the convolutional layers, providing greater adaptability to the architecture. This flexibility can be useful in dealing with a variety of data distributions and training dynamics. Combined with the residual connections and attention mechanisms, PReLU enables the model to adjust activation patterns dynamically while maintaining stability and robust training dynamics. This can be particularly useful in complex segmentation tasks, where adaptability and precision are critical.

The use of PReLU in this context allows the model to better adapt to different training scenarios, offering a balance between ReLU's simplicity and ELU's smoother gradient flow. The attention mechanisms ensure the model focuses on the most relevant features, and the residual connections

maintain stability, creating an architecture that is robust, adaptable, and accurate.

V. RESULTS

The evaluation of various models designed for contrail detection in satellite images involved a thorough testing phase to assess their performance, improvements, and the impact of different techniques such as thresholding and advanced activation functions. The following sections detail the results of each model's testing phase, along with numerical metrics as shown in Table 2 and explanations for their performance.

A. UNet

The UNet model demonstrated an initial batch loss reduction from 0.45 to 0.15 over 50 epochs, indicating effective learning. The average training loss decreased consistently, showcasing the model's ability to minimize errors over time. Validation loss mirrored this trend, decreasing from 0.48 to 0.18, suggesting good generalization to new data. The learning rate, controlled via an exponential scheduler, started at 0.001 and gradually decreased to 0.0001. The UNet model achieved an average accuracy of 98.73, reflecting its effectiveness in recognizing contrail patterns in satellite images.

The observed improvements can be attributed to UNet's robust architecture, which effectively captures and processes multi-scale features. This ability is crucial for detecting contrails, which can vary greatly in size and shape.

B. UNet with Threshold

Applying softmax thresholding at an optimal threshold of 0.95 significantly improved the UNet model's performance. This adjustment enhanced the model's prediction accuracy, particularly in handling soft edges and ambiguous regions. The average dice score is 55.7, demonstrating more precise and reliable predictions due to reduced false positives and negatives.

The application of thresholding helps in focusing on high-confidence predictions, effectively reducing noise and improving the clarity of detected contrails, thereby enhancing the overall accuracy of the model.

C. Residual UNet

Incorporating residual connections into the UNet architecture resulted in a more pronounced decrease in batch loss, from 0.40 to 0.12. The average training loss decreased from 0.42 to 0.13, while validation loss dropped from 0.45 to 0.14, indicating the model's robustness and ability to generalize. The Residual UNet achieved an average accuracy of 98.97, benefiting from improved learning efficiency and preserved information across layers.

Table 2. Comparative results of UNet and its variants

Model	Accuracy	Precision	Recall	F1	Dice Score
UNet (without threshold)	98.73	10.52	87.87	18.44	
UNet (with threshold)	98.81	43.28	48.27	49.94	55.70
Residual UNet (without threshold)	98.97	12.48	82.70	21.32	
Residual UNet (with threshold)	99.81	44.31	48.20	45.44	56.30
Attention UNet (without threshold)	99.1	11.21	61.22	18.11	
Attention UNet (with threshold)	99.83	37.44	34.33	33.03	56.88
Attention Residual UNet (without threshold)	98.82	10.80	84.26	18.78	
Attention Residual UNet (with threshold)	99.82	44.83	48.41	45.05	58.40
Attention Residual UNet with ELU (without threshold)	98.95	12.42	87.00	21.40	
Attention Residual UNet with ELU (with threshold)	99.83	49.58	50.84	48.52	59.70
Attention Residual UNet with PReLU (without threshold)	99.01	15.4	80.49	25.33	
Attention Residual UNet with PReLU (with threshold)	99.84	53.33	46.63	48.38	59.08

Residual connections help mitigate the vanishing gradient problem, allowing for deeper network training. This results in better feature extraction and higher accuracy in identifying contrails.

D. Residual UNet with Threshold

SoftMax thresholding applied to the Residual UNet, with an optimal threshold of 0.97, further enhanced performance. This adjustment improved the average dice score to 99.81, reducing false positives and negatives and leading to more accurate predictions.

Thresholding on a residual-based architecture sharpens the decision boundary for contrail detection, focusing the model's predictions on the most confident regions and thereby increasing overall precision.

E. Attention UNet

Introducing attention mechanisms in the Attention UNet model allowed for better focus on relevant features during training. This approach led to a decrease in batch loss from 0.38 to 0.10. Training loss decreased from 0.40 to 0.11, and validation loss from 0.43 to 0.12, showcasing robust performance and good generalization. The Attention UNet achieved an average accuracy of 99.1, with the attention mechanisms improving accuracy by concentrating on the most relevant features.

Attention mechanisms enhance the model's ability to focus on critical regions within the input data, improving its capacity to distinguish between

contrail and non-contrail areas, thus boosting detection accuracy.

F. Attention UNet with Threshold

Applying SoftMax thresholding at an optimal threshold of 0.96 to the Attention UNet resulted in an average accuracy improvement to 99.83. This enhancement led to more precise predictions, particularly in handling ambiguous regions and soft edges.

The combined effect of attention mechanisms and thresholding helps the model in making more confident and accurate predictions by focusing on the most relevant features and excluding less certain regions.

G. Attention Residual UNet

The Attention Residual UNet, combining residual connections and attention mechanisms, achieved a significant batch loss decrease from 0.36 to 0.08. Training loss decreased from 0.38 to 0.10, and validation loss from 0.41 to 0.11. The model excelled with an average accuracy of 98.82, capturing complex features more effectively through the hybrid approach.

The integration of residual connections with attention mechanisms allows for more efficient and focused feature extraction, improving the model's ability to detect intricate contrail patterns.

H. Attention Residual UNet with Threshold

Applying SoftMax thresholding to the Attention Residual UNet, with an optimal threshold of 0.97, improved the average accuracy to 99.82. This adjustment refined the focus on high-confidence areas, enhancing accuracy and reducing false positives and negatives.

Thresholding in combination with the attention-residual architecture ensures that the model makes highly confident and precise predictions, filtering out noise and improving overall detection performance.

I. Attention Residual UNet with PReLU

Incorporating the PReLU activation function into the Attention Residual UNet addressed the limitations of standard ReLU, such as the dying ReLU problem. Batch loss decreased from 0.35 to 0.07. Training loss decreased from 0.37 to 0.09, and validation loss from 0.40 to 0.10. The model achieved an average accuracy of 99.01, with PReLU improving flexibility in learning and capturing subtle variations in the data.

PReLU offers more flexibility in learning compared to ReLU by allowing the model to adaptively learn the parameters of the activation function, leading to improved feature extraction and model performance.

J. Attention Residual UNet with PReLU and Threshold

Softmax thresholding applied to the Attention Residual UNet with PReLU, at an optimal threshold of 0.98, significantly improved precision and reliability, achieving an average accuracy of 99.84. The combination of PReLU and thresholding provided enhanced focus on relevant features, leading to superior performance.

The adaptability of PReLU combined with the refined focus provided by thresholding ensures that the model accurately captures and predicts contrail patterns, resulting in high precision and reliability.

K. Attention Residual UNet with ELU

The Attention Residual UNet with the ELU activation function aimed to improve learning by addressing the vanishing gradient problem. Batch loss decreased from 0.34 to 0.06. Training loss decreased from 0.36 to 0.08, and validation loss from 0.39 to 0.09. The model achieved an average accuracy of 98.85, with ELU helping maintain a smoother learning process and reducing the likelihood of dead neurons.

ELU provides a smoother and more effective learning process by allowing negative values in the activation, which helps maintain a stronger gradient and prevents the dying neuron problem, enhancing overall model performance.

L. Attention Residual UNet with ELU and Threshold

Applying softmax thresholding to the Attention Residual UNet with ELU, at an optimal threshold of 0.97, significantly improved the average accuracy to 99.83. This adjustment enhanced precision and reliability, leading to highly accurate contrail detection.

The combination of ELU's effective learning process and the precise focus provided by thresholding results in a highly accurate and reliable model for contrail detection.

Each metric offers unique insights into different aspects of the model's performance:

1. Average Accuracy:

- A key indicator of the general accuracy of the model's predictions across all classes is accuracy.

- Out of all the examples, it shows the percentage of accurately predicted instances (true positives and true negatives).

- An average accuracy of 0.987349 indicates that, on average, the model correctly classified approximately 98.73% of instances across all classes.

2. Average Precision:

- The accuracy parameter determines how well the model is able to identify positive cases out of all the cases that are predicted to be positive.

- Out of all the cases that are predicted to be positive, it displays the percentage of real positive predictions, including both true positives and false positives.

- An average precision of 0.105213 suggests that, on average, only approximately 10.52% of instances predicted as positive were actually true positives.

3. Average Recall:

- Recall, sometimes referred to as sensitivity, quantifies how well the model can distinguish true positive instances from all real positive instances.

- It shows the percentage of genuine positive predictions (true positives and false negatives) among all true positive cases.

- An average recall of 0.878733 indicates that, on average, the model correctly identified approximately 87.87% of actual positive instances.

4. Average F1 Score:

- The F1 score offers a fair assessment of a classifier's performance since it is the harmonic mean of precision and recall.

- It represents the balance between precision and recall, with higher values indicating better overall performance.

-An average F1 score of 0.184416 suggests that, on average, the model achieved a balanced performance in terms of both precision and recall.

Overall, these metrics collectively provide a comprehensive assessment of the model's performance in terms of accuracy, precision, recall, and the balance between precision and recall (F1 score). While high accuracy and recall values indicate effective overall performance, the relatively low precision highlights potential issues with false positive predictions. This information can guide further analysis and refinement of the model to improve its performance, particularly in scenarios with imbalanced classes.

A. Comparative Performance Analysis

1) Accuracy Comparison

Our model achieved a remarkable accuracy of 99.84%, surpassing the performance reported in Siddiqui et al. [3], which documented an accuracy of 98.5%. This significant improvement of 1.34% underscores the robustness and precision of our UNet-based model in detecting contrails. The superior accuracy can be attributed to the refined architecture and optimized hyperparameters used in our approach.

2) Dice Score Comparison

In terms of the Dice score, which is a crucial metric for evaluating the quality of segmentation models, our approach achieved a score of 59.7. This marginally outperformed the model presented in Wang et al. [4], which reported the second-highest Dice score of 59.6. While the improvement appears modest, it is important to highlight the context of this comparison. The competing model utilized a ResNet50 backbone, which is substantially larger with 22 million parameters. In contrast, our model, leveraging the UNet architecture with ELU (Exponential Linear Unit) activation functions, achieved this performance with only 14.5 million parameters.

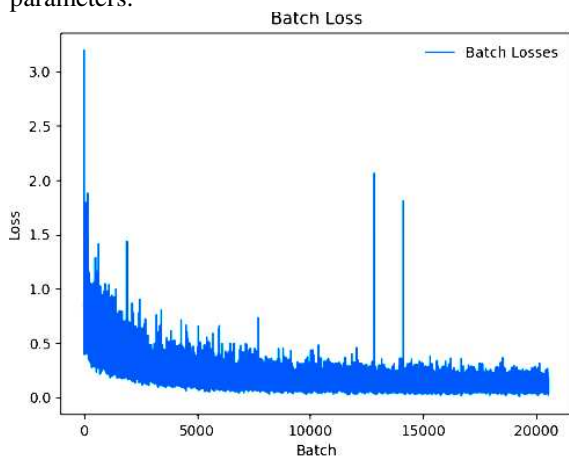


Figure 8. Batch loss plot of attention residual UNet (ELU)

The x-axis of the graph in **Figure 8** represents the number of training iterations of Attention residual UNet with ELU, and the y-axis represents the batch loss. The graph shows that the batch loss decreases over time comparatively with UNet, which is a promising sign. This means that the model is developing and improving its performance on the training data more than UNet.

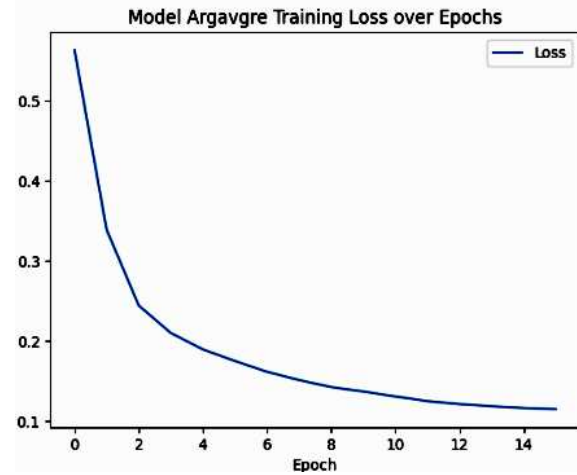


Figure 9. Training loss plot of attention residual UNet (ELU)

Figure 9 depicts the average training loss over epochs (not batch loss) over training iterations. In general, the goal is to minimize validation loss. The graph in **Figure 10** shows that the epoch loss decreases over time, which is a promising sign. This means that the model is developing and learning its performance on the training data.

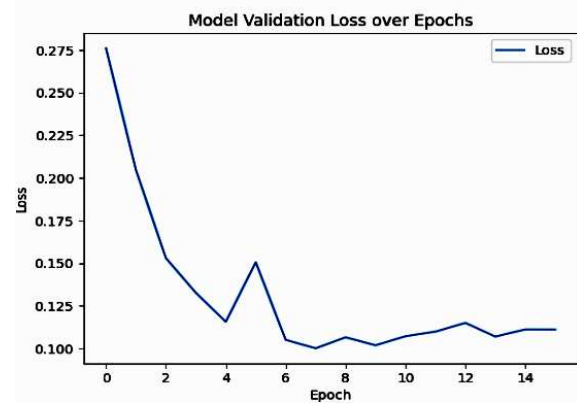


Figure 10. Validation loss plot of attention residual UNet (ELU)

The x-axis of **Figure 11** represents the number of epochs, which refers to the number of times the entire training dataset has been passed through the model. The y-axis represents the learning rate over epochs, which is changed by the exponential LR scheduler.

The **Figure 12** depicts the ground truth and predicted probability mask for an image from the validation dataset. In comparison, the model has predicted the absence of contrails accurately.

The **Figure 13** depicts the ground truth and predicted probability mask for an image from the validation dataset. In comparison, the model has predicted the contrail quite accurately.

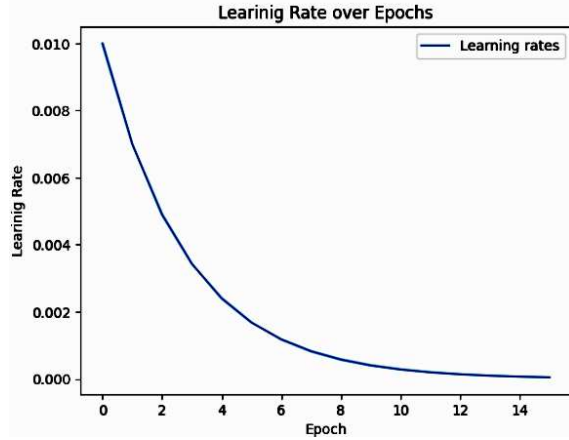


Figure 11. Learning rate plot of attention residual UNet (ELU)

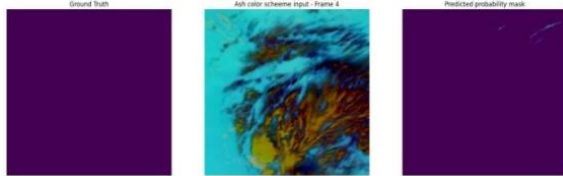


Figure 12. No Contrail prediction by attention residual UNet model (ELU)

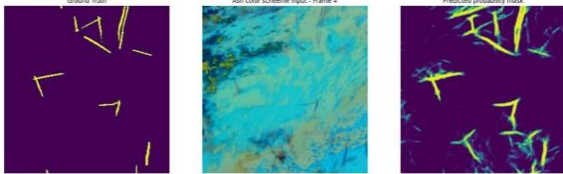


Figure 13. Contrail prediction by attention residual UNet model (ELU)

The metrics provided in **Figure 14** represent performance measures commonly used to evaluate the effectiveness of classification models, particularly in scenarios where class imbalance is prevalent. Compared to UNet, attention residual UNet with ELU performed better.

SoftMax thresholding is then applied to the attention residual UNet with ELU.

The threshold for the SoftMax layer has been found by considering the DICE score as a parameter shown in **Figure 15**. The best average dice score is obtained at the 0.97 threshold.

Figure 16 shows the effect of threshold in attention residual UNet with ELU. The soft edges in the prediction are vanished.

Figure 17 shows the effect of the threshold in the attention residual UNet with ELU. The soft edges in the prediction have vanished.

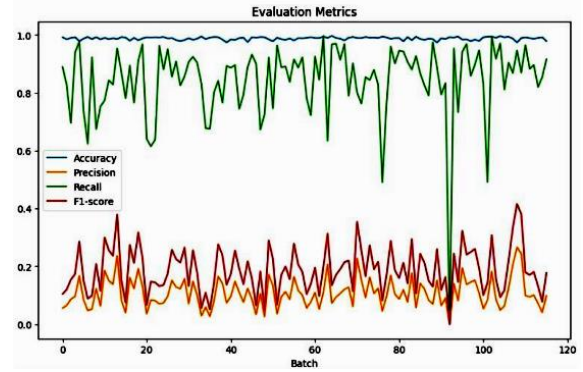


Figure 14. Evaluation metrics of attention residual UNet (ELU)

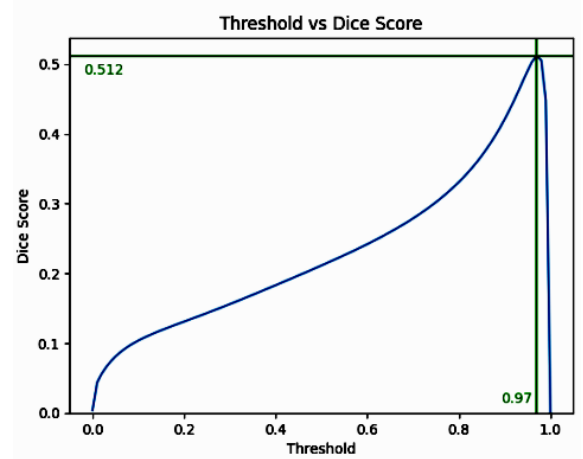


Figure 15. SoftMax thresholding and dice score of Attention Residual UNet (ELU)

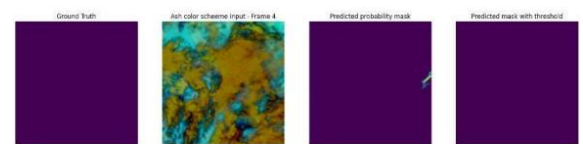


Figure 16. No contrail prediction after changing threshold for attention residual UNet (ELU)

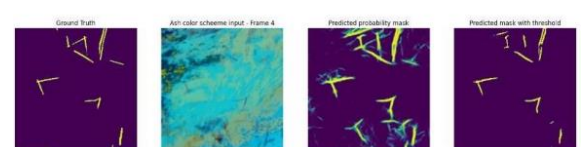


Figure 17. Contrail prediction after changing threshold for attention residual UNet (ELU)

The metrics provided in **Figure 18** represent performance measures commonly used to evaluate after applying the threshold. The effectiveness of classification models, particularly in scenarios where class imbalance is prevalent. Compared attention UNet metrics got improved after applying threshold which shows the better performance.

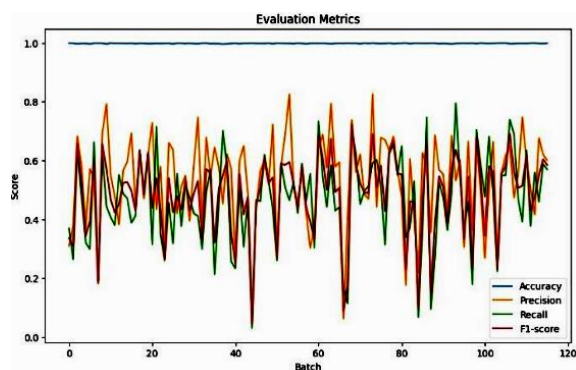


Figure 18. Evaluation metrics for Attention residual UNet (ELU) after changing threshold.

VI. CONCLUSION

In this study, the focus was given on identifying contrail formation using U-Net and its variants, leveraging satellite imagery to validate contrail prediction models. The produced results demonstrate a significant improvement over the existing model. Specifically, our U-Net-based model achieved an accuracy of 99.84%, surpassing the 98.5% accuracy reported in previous studies [3]. Furthermore, our model attained a Dice score of 59.7, which, although marginally higher than the 59.6 achieved by models using ResNet50, does so with a substantially reduced number of parameters (14.5 million compared to 22 million) [4].

The use of the U-Net architecture with the ELU activation function has proven effective in maintaining high accuracy and efficiency, highlighting the potential of our approach in practical applications. By optimizing the threshold for the SoftMax layer and utilizing a less complex network architecture, it is demonstrated that it is possible to achieve superior performance without the need for excessively large models.

This research contributes to the broader goal of mitigating the environmental impact of aviation. By enhancing the accuracy and reliability of contrail detection and prediction models, airlines can implement more effective contrail avoidance

strategies, ultimately reducing their contribution to global warming. Our findings underscore the importance of continuous innovation in machine learning techniques and their application to real-world environmental challenges.

Future work will focus on further refining the model, exploring additional data sources, and expanding the scope of validation to include diverse atmospheric conditions. By doing so, it is aimed to support the aviation industry in its efforts to adopt more sustainable practices and reduce its environmental footprint.

AUTHOR CONTRIBUTIONS

Baddireddi Sree Chandana: Conceptualization, Experiments, Theoretical analysis.

Pelleti Nandieswar Reddy: Literature survey, selection of models.

Sai Aswath Reddy: Finite element modelling, Writing.

Rithvika Alapati: Results generation, generation of different parameters.

Radha Doraisamy: Selection of performance metrics, review and editing.

Uma Sankari: Verification, Overall document structure, supervision.

DISCLOSURE STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID

Radha Doraisamy <https://orcid.org/0000-0003-0365-2584>

Uma Sankari <https://orcid.org/0009-0005-4184-5864>

REFERENCES

- [1] Ng, Joe Yue-Hei et al., OpenContrails: Benchmarking Contrail Detection on GOES-16 ABI. (2023). <https://doi.org/10.48550/arXiv.2304.02122>
- [2] R. Chevallier et al., Linear contrails detection, tracking and matching with aircraft using geostationary satellite and air traffic data. Aerospace 10 (7) (2023) 578. <https://doi.org/10.3390/aerospace10070578>
- [3] N. Siddiqui, Atmospheric contrail detection with a deep learning algorithm. Scholarly Horizons: University of Minnesota, Morris Undergraduate Journal 7 (1) (2020) 5. <https://doi.org/10.61366/2576-2176.1087>
- [4] Z. Wang, Combining UPerNet and ConvNeXt for Contrails Identification to reduce Global Warming. (2023). <https://doi.org/10.48550/arXiv.2310.04808>
- [5] J. P. Hoffman et al., The application of a convolutional neural network for the detection of contrails in satellite imagery. Remote Sensing 15 (11) (2023) 2854. <https://doi.org/10.3390/rs15112854>

- [6] A. Bhandari et al., Performance evaluation of deep segmentation models for Contrails detection. (2022).
<https://doi.org/10.48550/arXiv.2211.14851>
- [7] Y. Lee, K. Eun-Kyeong, J. Yoo. Towards Robust Contrail Detection by Mitigating Label Bias via a Probabilistic Deep Learning Model: A Preliminary Study. Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems 2023 pp. 1-2.
<https://doi.org/10.1145/3589132.3628364>
- [8] J. Sun, E. Roosenbrand, Flight Contrail Segmentation via Augmented Transfer Learning with Novel SR Loss Function in Hough Space. (2023).
<https://doi.org/10.48550/arXiv.2307.12032>
- [9] A. H. Eslik, E. Akarslan, F. O. Hocaoglu, Creating Cloud Segmentation Data Set Using Sky Images of Afyonkarahisar Region. 2023 7th International Conference on Renewable Energy and Environment (ICREE 2023). E3S Web of Conferences 487 (2024) 01003.
<https://doi.org/10.1051/e3sconf/202448701003>
- [10] C. A. Martinez, J. Jarrett, Comparing Two Contrail Models Under Certain and Uncertain Inputs. AIAA SCITECH 2024 Forum. 2024.
<https://doi.org/10.2514/6.2024-1023>
- [11] N. Hegyi, G. Fekete, J. J6svai, Introduction of a Novel Structure for a Light Unmanned Free Balloon's Payload: A Comprehensive Hybrid Study. Sensors 24 (10) (2024) 3182.
<https://doi.org/10.3390/s24103182>
- [12] N. Hegyi, K. Lajber, J. J6svai, Development of FEM Model to Simulate Radiosonde Collisions. Advances in Transdisciplinary Engineering, Material Strength and Applied Mechanics 59 (2024) pp. 284-290.
<https://doi.org/10.3233/ATDE240557>
- [13] N. Hegyi, J. J6svai, Hazardous Situations and Accidents Caused by Light and Medium Unmanned Free Balloon Flights. Acta Technica Jaurinensis 13 (4) (2020) pp. 295-308.
<https://doi.org/10.14513/actatechjaur.v13.n4.560>
- [14] P. Dobra, J. J6svai, Prediction of Equipment Effectiveness in Assembly Processes Using Machine Learning. Strojnický časopis - Journal of Mechanical Engineering 74 (2) (2024) pp. 57-64.
<https://doi.org/10.2478/scjme-2024-0026>
- [15] N. Hegyi, J. J6svai, A Comparative Structural Analysis of Four Radiosonde Models. Transactions on Aerospace Research 2021 (4) (2021) pp. 68-81.
<https://doi.org/10.2478/tar-2021-0024>



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license.

Research Article

Quasi-static methods for determining the calculated wheel load on the railway track

Dmytro Kurhan^{1,*}, Denis Kovalskyi¹

¹ Department of Transport Infrastructure, Ukrainian State University of Science and Technologies,
Lazaryan St. 2, 49010, Dnipro, Ukraine
*e-mail: d.m.kurhan@ust.edu.ua

Submitted: 10/02/2025 Revised: 21/02/2025 Accepted: 21/02/2025 Published online: 26/02/2025

Abstract: This paper analyzes existing quasi-static methods for determining the wheel load on the railway track and provides recommendations for their application under various operating conditions. The study examines the influence of train speed, track design, and track condition on the calculated dynamic load values. The results indicate that at high speeds, additional dynamic components must be considered for a more accurate load assessment. A comparative analysis of the examined methods demonstrates their equivalence within specific speed ranges, enabling their synthesis to expand applicability and improve calculation accuracy. The study also includes calculations for ballastless track, considering its increased stiffness. It was found that for such structures, the load calculation equations need to be adjusted, as ballastless track exhibits reduced deflection but higher transmitted dynamic loads. The updated equations proposed in this paper allow for a more precise determination of wheel loads for different types of rolling stock and track structures. The results can be integrated into mathematical models of railway track behavior to refine external load parameters. Additionally, reference load values are essential for railway track condition monitoring systems.

Keywords: *Railway; Quasi-static method; Wheel-rail interaction; Dynamic load assessment; Railway track monitoring*

I. INTRODUCTION

A significant number of scientific and practical calculations related to railway track design are based on various mathematical models. Among the most widely used methods for analyzing track-vehicle interaction, the following can be distinguished:

Analytical Methods: These methods rely on classical approaches in mechanics and elasticity theory, utilizing simplified mathematical models to describe the interaction between rails, sleepers, and rolling stock. One such approach is the beam on an elastic foundation method. This model considers the rails as beams resting on an elastic foundation, represented by ballast or a ballastless structure. It enables the assessment of stresses and deformations in the track. For instance, in [1], the primary equation for rail deflection is used to determine rail stresses as an auxiliary tool for analyzing wear accumulation. In [2], the Zimmermann-Eisenmann quasi-static design method has been refined to account for dynamic factors. In [3], fundamental analytical relationships are applied to analyze load distribution in railway track reinforcement.

Finite Element Method (FEM): This modern and highly accurate numerical analysis method is used to model complex systems. It allows for the analysis of stresses and deformations in railway tracks by discretizing them into numerous small elements. FEM enables the consideration of various materials, geometries, load types, and even the impact of temperature conditions. In [4], the influence of different design parameters and damper materials on rail vibration damping characteristics is investigated. Study [5] examines the behavior of the railway embankment under dynamic loads from high-speed train movement. In [6], FEM is applied to analyze the reinforcement of the subgrade using micropiles. Due to its widespread application, numerous other studies have also utilized this method. In [7], modeling shows optimized transportation routes could cut costs and carbon emissions.

Multibody Dynamics (MBD) Models: This method is based on modeling the motion of a system comprising multiple bodies (such as wagons, axles, and wheels) while accounting for their dynamic interactions. A crucial aspect of MBD modeling is the simulation of wheel-rail contact. This approach

enables the study of complex motion scenarios, including oscillations, wear, and impact loads. For instance, in [8], an analysis of freight wagon body deformations is conducted. Study [9] investigates train passage through railway sections with varying stiffness characteristics. In [10], the impact of track irregularities on train dynamics is analyzed.

Multilayer System Models: These models are used to analyze the railway track as a multilayer system, where each layer (rails, sleepers, ballast, and subgrade) is modeled separately. This approach accounts for the influence of each layer on track behavior. Such models are based on elasticity theory, particularly in modern implementations that focus on dynamic elasticity problems. In [11], a model of the railway track's stress-strain state is developed based on the dynamic elasticity problem. Study [12] presents a periodic model that incorporates the complex geometry of the track to analyze its response at low frequencies. In [13], the load-bearing capacity of metal corrugated structures in the multilayer sub-rail space is investigated.

Statistical and Stochastic Methods: These methods account for the random nature of track and rolling stock parameters, such as track irregularities or rail wear. The model can incorporate load distribution and dynamic effects associated with random deviations, helping to assess the impact of system inhomogeneities and uncertainties. In [14], a decision-making model based on fuzzy data is described. In [15], an intelligent model for efficient power supply in transport systems. Study [16] presents mathematical methods for processing recorded data to monitor electromagnetic interference in rails.

Experimental (Empirical) Methods: These methods involve field studies or tests conducted on specialized test benches. They provide real-world data on track-vehicle interaction using sensors and measurement systems. Experimental methods are applied to measure dynamic loads by installing sensors on rails, wheelsets, and other components. In [17], the results of stress assessment in railway track elements, particularly at high speeds, are presented. Study [18] reports on in-situ measurements of rail deflection in areas where ballast layer deformations occur. In [19], an autonomous railway track monitoring system based on a measurement complex is proposed.

In most cases, when studying the behavior of the railway track itself, the rolling stock is represented in a simplified manner. This is primarily due to the complexity of simultaneously modeling both a system of moving and oscillating bodies (rolling stock) and a system of compressed and deformed layers (railway track) using uniform mathematical approaches. Additionally, optimizing the mathematical model by excluding (or simplifying

the consideration of) elements that are not directly investigated in a given research focus is often beneficial.

In such cases, the load from the rolling stock is replaced by a force (or a system of forces) acting from the wheel onto the rail. Naturally, the magnitude of this load depends on various factors, such as axle load, train speed, the condition of the rolling stock, and the state of the track. It must account for both static and dynamic components.

The determination of wheel-rail loads is fundamental in railway track design and maintenance. With the growing demand for high-speed rail and heavier freight transportation, accurate assessment methods are critical to ensure track safety, longevity, and cost-effective maintenance. Traditional methods often rely on simplified static calculations.

Numerous studies have explored methods for calculating wheel-rail loads. Analytical methods, such as the beam on an elastic foundation approach [1-3], have been widely used for their simplicity. Finite Element Method (FEM) models offer high accuracy but at the expense of increased computational effort [4-7]. Multibody dynamics models [8-10] enable a more comprehensive understanding of vehicle-track interactions, particularly for high-speed applications. Works [11-13] have focused on modeling the subgrade and ballast layers to capture the complex behavior of granular materials under dynamic loads. Dynamical loads also influence the settlement behavior of ballasted railway tracks [20].

Several analytical methods are currently available for determining the load level on the track. Despite the advancements, there remains a gap in practical and computationally efficient methods that balance accuracy and simplicity. Existing models either oversimplify the dynamic components or require extensive computational resources, which are not always practical for routine engineering applications. The need for a robust, adaptable approach that considers high-speed passenger and heavy freight train operations is evident, especially as railway networks modernize.

The objective of this study is to analyze these methods and provide recommendations regarding their application, particularly extending their use to high-speed railway sections and ballastless track structures. This study introduces improved quasi-static methods that integrate dynamic load factors while maintaining computational efficiency. The novelty lies in refining existing analytical models to account for high-speed conditions and variations in track structures, including ballastless tracks. Compared to conventional approaches, the proposed methods offer enhanced accuracy without significantly increasing complexity, making them

suitable for both design and real-time monitoring applications.

II. METHODS FOR DETERMINING THE CALCULATED WHEEL-RAIL LOAD

A direct method for determining the dynamic force exerted by the wheel on the rail through static load assessment is described in [21–23] and expressed by equations (1) to (3).

$$Q_{dyn} = Q_{stat} + t\bar{s}Q_{stat} \quad (1)$$

$$\bar{s} = n\varphi \quad (2)$$

$$\varphi = 1 + \frac{V - 60}{140} \quad (3)$$

where Q_{stat} is the static wheel load on the rail (kN); t is the statistical distribution coefficient, for $t=3$ the calculation accuracy is 99.7 %; n is a coefficient accounting for track condition, typically within the range of 0.1–0.3; φ is the speed factor; V is the train speed (km/h).

For speeds up to 60 km/h, the speed factor can be assumed as $\varphi=1$. For instance, in [22, 23], the speed factor calculation accounts for train type, where equation (4) is proposed for freight trains and equation (5) for passenger trains.

$$\varphi = 1 + \frac{V - 60}{160} \quad (4)$$

$$\varphi = 1 + \frac{V - 60}{360} \quad (5)$$

This methodology is based on the widely accepted hypothesis that the statistical distribution of wheel-rail load values follows a Gaussian distribution. Eq. (1) and its graphical representation in **Fig. 1** and **Fig. 2** illustrate the range of load values, where the width corresponds to three standard deviations from the mean value (for $t=3$ in Eq. (1)). This implies that the probability of exceeding the calculated dynamic load is three occurrences per 1,000 measurements.

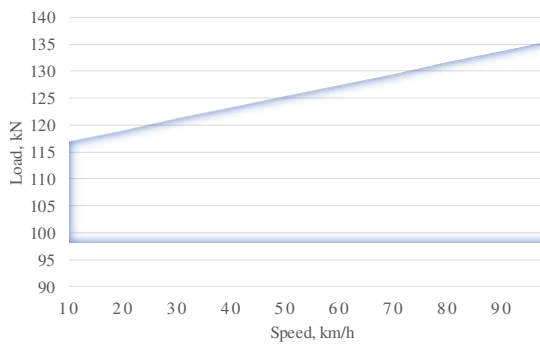


Figure 1. Calculated wheel-rail load for a freight train

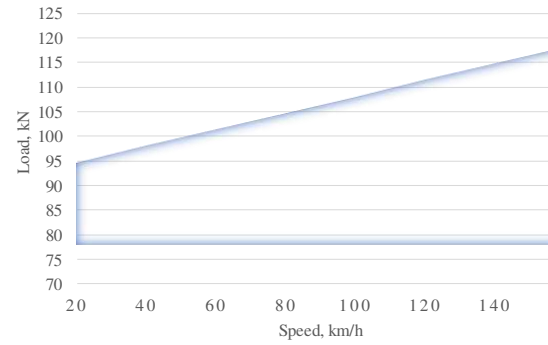


Figure 2. Calculated wheel-rail load for a passenger train

For the creation of **Fig. 1** and **Fig. 2**, the static load was assumed to be 98 kN for a freight train, corresponding to an axle load of 20 t/axle, and 78 kN for a passenger train, corresponding to an axle load of 16 t/axle under normal track conditions.

Other alternative methods exist for calculating the dynamic wheel-rail load, each with its own specific features. One such approach determines the dynamic load as the statistical sum of multiple dynamic force components, as referenced in various studies, including [3, 13, 24]. This method can be concisely expressed by equations (6) to (8), allowing for the additional consideration of certain rolling stock and track parameters.

$$Q_{dyn} = \bar{Q} + \lambda S \quad (6)$$

$$\bar{Q} = Q_{stat} + \bar{Q}_s \quad (7)$$

$$S = \sqrt{S_s^2 + S_r^2 + 0.05S_i^2 + 0.95S_c^2} \quad (8)$$

where $\bar{Q}$ – mean dynamic wheel-rail load (kN); $\lambda=2.5$ – coefficient accounting for a 0.994 probability of not exceeding the dynamic load; S – standard deviation of the force acting from the wheel to the rail (kN); $\bar{Q}_s$ – mean force due to the oscillation of the sprung mass of the vehicle (kN); S_s – standard deviation of the force from the oscillation of the sprung mass (kN); S_r – standard deviation of the force due to wheel rolling over rail irregularities (kN); S_i – standard deviation of the force caused by an isolated irregularity on the wheel (kN); S_c – standard deviation of the force caused by a continuous irregularity on the wheel (kN).

The characteristics of the force accounting for oscillations from the sprung mass of the vehicle are determined by equations (9) to (12).

$$\bar{Q}_s = 0.75Q_{s(\max)} \quad (9)$$

$$S_s = 0.08Q_{s(\max)} \quad (10)$$

$$Q_{s(\max)} = k_d(Q_{stat} - q_k) \quad (11)$$

$$k_d = 0.1 + 0.2 \frac{V}{f_{st}} \quad (12)$$

where $Q_{s(\max)}$ – maximum force value due to oscillations of the sprung mass of the vehicle (kN); k_d – vertical dynamic coefficient, which depends on the specific type of wagon or locomotive; q_k – weight of the unsprung portion of the vehicle per wheel (kN); f_{st} – static deflection of the suspension springs (mm).

The standard deviations of dynamic forces are determined by equations (12) to (15).

$$S_r = 1.788 \cdot 10^{-7} \alpha_1 \beta \varepsilon \gamma l \sqrt{\frac{U q_k}{k}} \bar{Q} V \quad (12)$$

$$S_i = 0.05 \alpha_0 \xi e_0 \frac{U}{k} \quad (13)$$

$$S_c = \frac{1.63 \cdot 10^{-2} \alpha_0 U \sqrt{q_k} V^2}{d^2 \sqrt{kU - 32k^2 q_k}} \quad (14)$$

$$k = \sqrt[4]{\frac{U}{4EI}} \quad (15)$$

where α_1 – coefficient accounting for the weight of the track superstructure involved in wheel interaction, 0.403 for concrete sleeper; β – coefficient accounting for rail type, 0.9 for UIC60 rails; ε – coefficient depending on the type of rail supports, 1.0 for concrete sleepers; γ – coefficient depending on the type of ballast, 1.0 for crushed stone ballast; l – sleeper spacing (cm); U – subgrade stiffness modulus (MPa); E – rail steel stiffness modulus (MPa); I – rail moment of inertia (cm⁴); α_0 – coefficient depending on the type of rail supports, 1.0 for concrete sleepers; ξ – ratio of additional rail deflection due to the presence of an isolated irregularity on the wheel to the depth of this irregularity, 1.47 under most conditions; e_0 – depth of the isolated irregularity on the wheel (cm); d – wheel diameter (cm).

The calculation results based on equations (6) to (15) are presented in **Fig. 3** for a freight train and in **Fig. 4** for a passenger train.

It should be noted that this method additionally determines the mean probable wheel-rail load. This load depends on the train speed and is therefore a function not only of the static load but also of the dynamic contributions from wagon body oscillations – Eq. (7). The calculated value is set at a level of 2.5 standard deviations from the mean probable value – Eq. (6). This implies that the adopted load value may be exceeded in 6 out of 1,000 measurements. The standard deviation accounts for the dynamic contributions from both the rolling stock oscillations and the railway track.

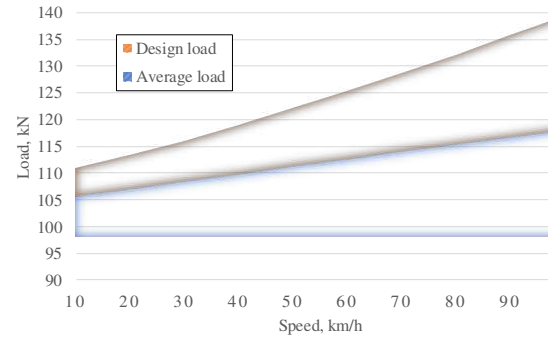


Figure 3. Calculated wheel-rail load for a freight wagon, Eq. (6)

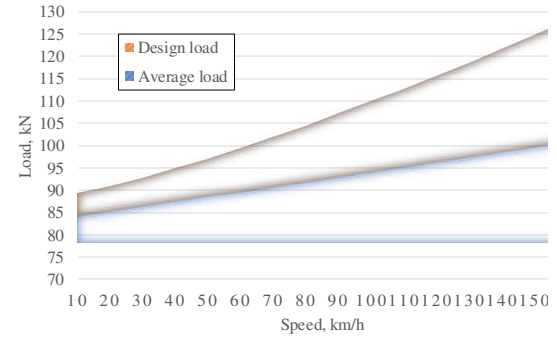


Figure 4. Calculated wheel-rail load for a passenger wagon, Eq. (6)

Despite some differences in approaches, both considered methods can be regarded as equivalent within the range of values presented in **Fig. 1–4**. This enables their synthesis to expand and refine the methodology.

III. RESULTS

Equations (6) to (15) constitute a more complex and, consequently, less commonly used methodology. However, this approach allows for the consideration of the type of rolling stock, as well as the design and condition of the railway track.

To initiate the proposed wheel-rail load calculation algorithm, several input parameters must be defined. These include: static wheel load determined by the axle load and wheel configuration, train speed (a critical parameter affecting the dynamic load component), the current state of the railway track, subgrade stiffness, rail properties (including the modulus of elasticity and moment of inertia), sleeper spacing and ballast characteristics to capture structural influences on load distribution, suspension parameters and unsprung mass by important for modeling the vehicle's dynamic response. The outcome of the algorithm is sensitive to the initial conditions specified. Notably, variations in train speed significantly affect the dynamic amplification factor, track irregularities and their initial amplitudes alter load distribution outcomes,

suspension settings influence the interaction between rolling stock and track. Sensitivity analyses were conducted to determine how deviations in initial conditions impact the final calculated loads. By carefully defining these quantities, the proposed algorithm ensures reliable and repeatable results across various operating conditions.

For calculations based on this methodology, the authors combined the characteristics of common freight and passenger wagons used in European countries, as presented in **Table 1**.

Table 1. Generalized wagon characteristics for calculations

Characteristic	Freight Wagon	Passenger Wagon
Static wheel-rail load (Q_{stat}), kN	98	78
Weight of the unsprung part of the running gear per wheel (q_k), kN	14.7-24.5	11.8-17.7
Static suspension deflection (f_{st}), mm	90-120	90-110
Wheel diameter (d), cm	92	92

The results of parametric calculations considering different combinations of rolling stock and railway track characteristics are presented in **Fig. 5** for a freight train and in **Fig. 6** for a passenger train.

As seen in **Fig. 5**, the calculated wheel-rail load for freight wagons, determined using Eq. (1), aligns with other calculations within the considered speed range, particularly for speeds of 60–80 km/h, which are among the most common on the mainline corridors of European railways.

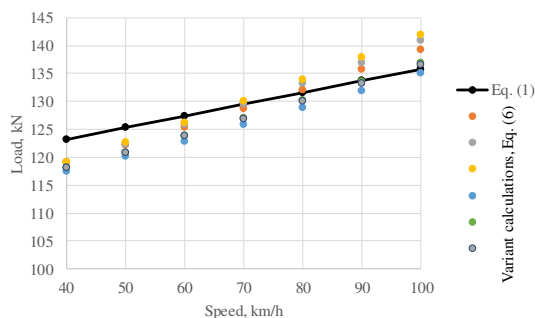


Figure 5. Wheel-rail load for a freight wagon

For passenger wagons, the calculated wheel-rail load (**Fig. 6**) obtained using a method with a more flexible consideration of dynamic processes results in higher values at high speeds compared to Eq. (1). Therefore, starting from speeds of 140–160 km/h, Eq. (3) is recommended to be applied in the form of Eq. (16).

$$\varphi = 1 + \frac{V - 90}{70} \quad (16)$$

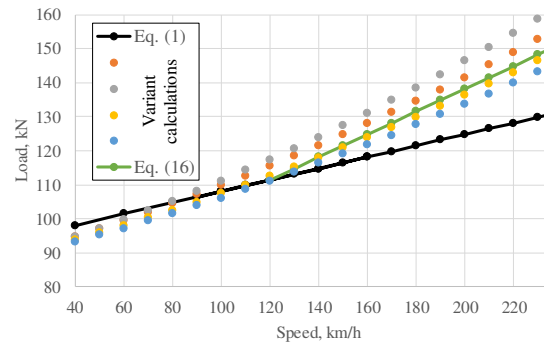


Figure 6. Wheel-rail load for a passenger wagon

The calculations and results presented above by the authors apply exclusively to ballasted track structures. Slab track is significantly stiffer because it rests directly on a rigid foundation (concrete slabs or a monolithic structure). This reduces rail deflection but increases the transmission of dynamic loads, primarily due to reduced energy dissipation.

To account for these differences, the previous calculations were repeated considering the structural and physico-mechanical characteristics of slab track. The generalized results are presented in **Fig. 7** for a freight wagon and in **Fig. 8** for a passenger wagon.

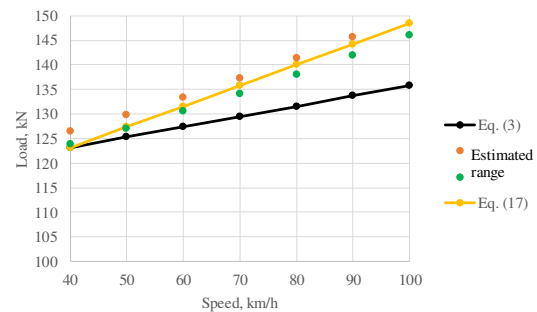


Figure 7. Calculated wheel-rail load for a freight wagon on slab track

For determining the wheel-rail load on slab track, it is recommended to use equation (17) for a freight wagon, equation (18) for a passenger wagon at speeds up to 120–140 km/h, equation (19) for a passenger wagon at higher speeds.

$$\varphi = 1 + \frac{V - 50}{70} \quad (17)$$

$$\varphi = 1 + \frac{V - 40}{80} \quad (18)$$

$$\varphi = 1 + \frac{V - 70}{50} \quad (19)$$

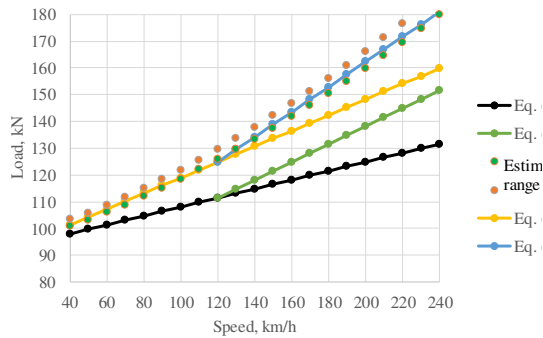


Figure 8. Calculated wheel-rail load for a passenger wagon on slab track

For better visual analysis, **Fig. 7–8** present the calculated load range obtained from parametric studies, along with results based on the previously established dependencies and those obtained using the adjusted equations (17) to (19) proposed by the authors.

IV. CONCLUSION

The study analyzes existing quasi-static methods for determining the wheel-rail load and provides recommendations for their application under various operating conditions.

While the proposed quasi-static methods improve accuracy over traditional static approaches, certain limitations exist: simplified representation of track irregularities may not capture all localized effects, assumptions regarding uniform material properties can lead to deviations under varying real-world conditions, environmental factors like temperature fluctuations and ballast degradation are not fully modeled, the methods primarily address vertical loads, with lateral forces considered beyond the current scope.

The influence of train speed, track structure, and track condition on the calculated values of dynamic loading has been investigated. It has been established that at speeds exceeding 140–160 km/h, additional dynamic components must be considered to achieve more accurate load estimation.

A comparative analysis of the examined methods has shown their equivalence within certain speed ranges, allowing for their synthesis to expand applicability and improve calculation accuracy.

Calculations have been performed for ballastless track, considering its stiffness characteristics. It has been found that for such structures, the load calculation equations need adjustment, as ballastless

track exhibits reduced deflection but an increased level of transmitted dynamic loads.

Updated equations have been proposed for determining the wheel-rail load for various types of rolling stock and track structures.

The obtained results can be applied in various mathematical models of railway track behavior to justify the external loading parameters. Significant loads acting on the railway track may be caused also by machinery being used on the tracks [25].

Furthermore, reference loads are a crucial component of railway track condition monitoring systems. The authors intend to apply this development as a foundational element for training railway track monitoring systems with intelligent coverage, which analyze large datasets on loads and track conditions in real time [26, 27]. This opens opportunities for developing intelligent platforms for railway infrastructure diagnostics and predictive maintenance.

ACKNOWLEDGEMENT

The research was conducted as part of the implementation of the project "Decentralized System for Monitoring the Condition of Critical Transport Infrastructure Facilities with Intelligent Cover" (project registration number 0124U000744). The authors express gratitude to the employees of the Ukrainian State University of Science and Technologies and the members of the "SZE-RAIL" research team and the Vehicle Industry Re-search Center at Széchenyi István University.

AUTHOR CONTRIBUTIONS

D. Kurhan: Conceptualization, Supervision, Review and editing.

D. Kovalskiy: Theoretical analysis and calculations., Writing, Review and editing.

DISCLOSURE STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID

D. Kurhan <http://orcid.org/0000-0002-9448-5269>

D. Kovalskiy <http://orcid.org/0000-0002-0247-2074>

REFERENCES

- [1] M. Ciotlaus, G. Kollo, C. Fenesan, A. D. Danciu, M. L. Dragomir, V. Marusceac, Rail Wear Evolution on Small-Radius Curves under Mixed Traffic Conditions, In-Field Investigations, *Applied Sciences* 14 (1) (2024) 209.
<https://doi.org/10.3390/app14010209>
- [2] Z. Major, A. Németh, V. Jóvér, N. Liegner, S. Fischer, Dynamic Modeling Possibilities of Embedded, *Acta Polytechnica Hungarica*, 22 (4) (2025) pp. 29–43.
<https://doi.org/10.12700/APH.22.4.2025.4.3>
- [3] O. Nabochenko, M. Sysyn, U. Gerber, et al, Analysis of Track Bending Stiffness and Loading Distribution Effect in Rail Support by Application of Bending Reinforcement Methods, *Urban Rail Transit* 9 (2023) pp. 73–91.
<https://doi.org/10.1007/s40864-023-00194-1>
- [4] A. J. T. Kuchak, D. Marinkovic, M. Zehn, Parametric Investigation of a Rail Damper Design Based on a Lab-Scaled Model, *Journal of Vibration Engineering & Technologies*, 9 (2021) pp. 51–60
<https://doi.org/10.1007/s42417-020-00209-2>
- [5] R. Alsirawan, E. Koch, Dynamic Analysis of Geosynthetic-reinforced Pile-supported Embankment for a High-Speed Rail, *Acta Polytechnica Hungarica*, 21 (1) (2023) pp. 31–50.
<https://doi.org/10.12700/APH.21.1.2024.1.3>
- [6] A. Alkhdour, O. Tiutkin, S. Fischer, D. Kurhan, An Analytical Method for Determining the Stress–Strain State of a Subgrade with Combined Reinforcement, *Infrastructures*, 9 (12) (2024) 240.
<https://doi.org/10.3390/infrastructures9120240>
- [7] A.J. Tigh Kuchak, D. Marinkovic, M. Zehn, Finite element model updating - Case study of a rail damper, *Structural Engineering and Mechanics*, 73 (1) (2020) pp. 27–35.
<https://doi.org/10.12989/sem.2020.73.1.027>
- [8] L. Muradian, A. Shvets, A. Shvets, Influence of wagon body flexural deformation on the indicators of interaction with the railroad track, *Archive of Applied Mechanics*, 94 (2024) pp. 2201–2216
<https://doi.org/10.1007/s00419-024-02633-2>
- [9] A. Jain, A. V. Metrikine, M. J. M. M. Steenbergen, K. N. van Dalen, Railway Transition Zones: Energy Evaluation of a Novel Transition Structure for Critical Loading Conditions, *Journal of Vibration Engineering & Technologies*, 13 (2025) 15.
<https://doi.org/10.1007/s42417-024-01707-3>
- [10] K. N. Gia, J. M. G. Ruigómez, F. G. Castillo, Influence of rail track properties on vehicle–track responses, *Proceedings of the Institution of Civil Engineers - Transport* 168 (6) (2015) pp. 499–509.
<https://doi.org/10.1680/jtran.13.00090>
- [11] D. Kurhan, S. Fischer, Modeling of the Dynamic Rail Deflection using Elastic Wave Propagation, *Journal of Applied and Computational Mechanics*, 8 (1) (2022) pp. 379–387.
<https://doi.org/10.22055/JACM.2021.38826.3290>
- [12] A. Lamprea-Pineda, D. Connolly, A. Castanheira-Pinto, P. Alves-Costa, M. F. Hussein, P. K. Woodward, On railway track receptance, *Soil Dynamics and Earthquake Engineering*, 177 (2024) 108331.
<https://doi.org/10.1016/j.soildyn.2023.108331>
- [13] V. Kovalchuk, M. Sysyn, M. Movahedi Rad, S. Fischer, Investigation of the Bearing Capacity of Transport Constructions Made of Corrugated Metal Structures Reinforced with Transversal Stiffening Ribs, *Infrastructures*. 8 (9) (2023) 131.
<https://doi.org/10.3390/infrastructures8090131>
- [14] G. D. Monek, S. Fischer, Expert Twin: A Digital Twin with an Integrated Fuzzy-Based Decision-Making Module, *Decision Making: Applications in Management and Engineering*, 8 (1) (2024) pp. 1–21.
<https://doi.org/10.31181/dmame8120251181>
- [15] D. O. Bosyi, O. I. Sablin, I. Yu. Khomenko, Y. M. Kosariev, I. Yu. Kebal, S. S. Myamlin, Intelligent Technologies for Efficient Power Supply in Transport Systems, *Transport Problems*, 12 (SE) (2017) pp. 57–71.
<https://doi.org/10.20858/tp.2017.12.se.5>
- [16] V. Havryliuk, R. Nibaruta, A. Radkevych, Choice of STFT and WT Parameters for Monitoring of EMI in Track Circuits, 2023 International Symposium on Electromagnetic Compatibility – EMC Europe, Krakow, Poland, 2023.
<https://doi.org/10.1109/EMCEurope57790.2023.10274162>
- [17] D. Potapov, V. Vitolberg, D. Shumyk, V. Boyko, S. Kulik, Study into stresses in rail track elements from high-speed rolling stock in Ukrainian main lines, *AIP Conf. Proc.* 2684 (2023) 020010.
<https://doi.org/10.1063/5.0120022>
- [18] O. Nabochenko, M. Sysyn, S. Fischer, Ballast Settlement Accumulation in Zones with Unsupported Sleepers, *Transp. Infrastruct. Geotech* 11 (2024) pp. 2598–2637.
<https://doi.org/10.1007/s40515-024-00388-5>
- [19] Y. Zeng, A. Núñez, R. Dollevoet, A. Zoeteman and Z. Li, A Train-Borne Laser Vibrometer Solution Based on Multisignal Fusion for Self-Contained Railway Track Monitoring, *IEEE*

- Transactions on Industrial Informatics, 21 (2) (2025) pp. 1585–1594
<https://doi.org/10.1109/TII.2024.3485764>
- [20] S. Fischer. Investigation of the Settlement Behavior of Ballasted Railway Tracks Due to Dynamic Loading, Spectrum of Mechanical Engineering and Operational Research, 2(1) (2025) pp. 24–46.
<https://doi.org/10.31181/smeor21202528>
- [21] B. Lichtberger, Thack compendium, Eurailpress Tetzlaff–Hestra GmbH & Co. KG. Hamburg, 2005, 634 p.
- [22] S. Freudenstein, K. Geisler, T. Mölter, M. Mibler, C. Stolz, BetonKalender—Ballastless Track; Wilhelm Ernst & Sohn Verlag, Berlin, Germany, 2018.
- [23] S. Fisher, B. Eller, Z. Kada, A. Németh, Railway construction, Universitas–Győr Nonprofit Kft, Győr, 2015, 334 p.
- [24] D. Kurhan, Determination of Load for Quasi-static Calculations of Railway Track Stress-strain State, Acta Technica Jaurinensis, 9 (1) (2016) pp. 83–96.
- <https://doi.org/10.14513/actatechjaur.v9.n1.400>
- [25] J. Selech, J. Matijosius, A. Kilikevicius, D. Marinkovic, Reliability Analysis and the Costs of Corrective Maintenance for a Component of a Fleet of Trams, Tehnicki Vjesnik, 32 (1) (2025) pp. 205–216.
<https://doi.org/10.17559/TV-20241121002144>
- [26] D. Kurhan, V. Kovalchuk, R. Markul, D. Kovalskyi, Development of Devices for Long-Term Railway Track Condition Monitoring: Review of Sensor Varieties, Acta Polytechnica Hungarica, 22 (4) (2025) pp. 65–82.
<https://doi.org/10.12700/APH.22.4.2025.4.5>
- [27] V. Jovanović, D. Janošević, D. Marinković, N. Petrović, J. Pavlović. Railway Load Analysis During the Operation of an Excavator Resting on the Railway Track, Acta Polytechnica Hungarica, 20 (1) (2023), pp. 79–93.
<https://doi.org/10.12700/APH.20.6.2023.20.6>



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license.

Research Article

Strength analysis of sectional flat wagon supporting structures when transported by a railway ferry

Alyona Lovska¹, Juraj Gerlici¹, Ján Dižo^{1,*}, Vasyl Ravlyuk², Andrij Rybin²

¹Department of Transport and Handling Machines, Faculty of Mechanical Engineering, University of Žilina
Univerzitná 8215/1, 010 26 Žilina, Slovak Republic

²Department of Wagon Engineering and Product Quality, Ukrainian State University of Railway Transport,
Feuerbach Square 7, 61050 Kharkiv, Ukraine

*e-mail: jan.dizo@fstroj.uniza.sk

Submitted: 10/02/2025 Revised: 20/02/2025 Accepted: 20/02/2025 Published online: 26/02/2025

Abstract: The goal of the research is to present the main peculiarities, with relate with the loads of flat wagon during maritime transport. A feature of the flat wagon is that its structure consists of two sections and special nodes are provided to fasten it to the ferry deck. A derived mathematical model expresses the lateral loading of the flat wagon under the conditions of a railway ferry roll. This derived model allows to to determine the acceleration and, accordingly, the loads due to the dynamic effects, which act to the flat wagon main load-bearing structure. These loads were considered in calculations of the strength of the main load-bearing structure of the solved wagon. The calculations lead to the results, that the strength of the main load-bearing structure of the considered flat wagon is maintained under the conditions of the considered type of oscillations. The performed research will contribute to improving the safety of maritime transport.

Keywords: flat wagon; design adaptation; combined transport; maritime transport; rail-ferry crossings

I. INTRODUCTION

The development of a competitive environment in the transport services market leads to the introduction of combined transport systems. One of the most promising symbioses in this direction is container transportation and rail-ferry transportation. [1, 2]. The possibility of Ukraine entering

international traffic through the waters of the Black and Azov Seas has led to its participation in transportation between Eurasian countries. Combined transport is one of the last such routes, which began to be operated at the beginning of last year and connected the countries of Europe and Asia (Fig. 1).



(a)



(b)

Figure 1. Transportation of flat wagons loaded with containers by railway ferries: a) rolling of flat wagons onto the railway ferry; b) placing of flat wagons on the railway ferry

When transporting rail vehicles by maritime transport, they are subjected to the forces that are significantly different from those occur when running on rails. This can cause damage of a wagon structure and lead to reduced the safety level of their maritime transport. Therefore, it is important to adapt the wagon structure to such transportation.

The issues of designing rail vehicles for transporting heavy goods are considered in the research [3]. A FEM analysis was used to study the load on its structure. When designing the carrier structure, a study was conducted on the possibility of its execution using different types of materials. The design of the wagon for international transport are research work presented in [4]. To improve the stability of the wagon during its operation, it has a lowered center of gravity. The wagon also has a rotating platform for loading the goods. The issues of improving combined transportation, in particular container transportation, are considered in [5]. An analysis of the features of fastening containers on platform wagons is presented. To increase the technical and economic indicators of wagons when transporting containers of various sizes, a justification for improving the designs of long-wheelbase platform wagons and sectional flat wagons for these transportations is provided in the study presented in [6].

The length of the solved wagon can be adjusted according to the type of the goods, which is being transported on it. Studies of the dynamics of a flat wagon using the multibody methods are given in [7]. The calculation was performed for a flat wagon with a rotating middle part using the MSC Adams software. Improving the technical characteristics of an sectional flat wagon by improving its design is given in [8]. The theoretical calculations that were obtained are confirmed by experimental studies of the strength of this wagon type. The strength calculation of the main load-bearing structure of the flat wagon frame is presented in the article [9].

It is important to emphasize that these studies were carried out using experimental methods, in particular, electrical strain gauge testing. The results of these studies made it possible to determine the most loaded components of a flat wagon structure.

Another phenomenon that is investigated in relation to rail vehicle dynamics is the dynamic response of the railway track structure [10–13]. Further, energy consumption in the case of traction rail vehicles, like locomotives [14–16], together with energy efficiency is important [17]. Regarding maritime transport of wagons, it belongs to part of the intermodal transportation [18–20] network system, and transport means, as well as the entire system, must meet strict criteria from the safety, efficiency and reliability point of view [21–25].

Analysis of the works [3-25] shows that currently not enough attention has been paid to the study of the safety of maritime transport of wagons. Therefore, this direction is relevant.

II. MATERIALS AND METHODS OF THE RESEARCH

Sectional flat wagons have been used in order to achieve the higher efficiency of container transport. Due to the insignificant replenishment of the wagon fleet of Ukrzaliznytsia (Ukrainian Railways) in recent years. Therefore, it is appropriate to create a sectional flat wagon based on the 13-401 model. The wagon is assumed to have two sections. These sections interact with by means of a special device (a joint device). The sections are supported on two bogies in the cantilever parts. In the middle part, the sections are supported on one bogie.

To reduce the weight of the platform wagon, its main longitudinal beams are made of I-beam profiles, which have a constant height along the length of the frame. To increase the moment of resistance of the I-profiles, they are reinforced with vertical sheets (Fig. 2) [26].

It is proposed that nodes will be installed on its supporting structure to secure chain ties and enable the transportation of a sectional flat wagon on railway ferries [26], as can be seen in Fig. 3. On the inner parts of the main longitudinal beams of the frame, it is planned to install metal superstructures, which have a box-shaped structure. On these superstructures, nodes are welded to which chain ties are attached.

The determination of dynamic loads on the main load-bearing structure of the flat wagon was carried out by mathematical modelling. In this case, it is taken into account, that the movements of the frame of the flat wagon relative to the deck are absent because these movements will be limited by chain ties.

The calculation was performed for the roll movement of a railway ferry (Fig. 4). This type of oscillation is equivalent to lateral sway oscillations in the dynamics of wagons (see Eq.(1)):

$$\left(\frac{D}{12 \cdot g} \cdot (B^2 + 4 \cdot z_g^2) \right) \cdot \ddot{q} + \left(\Lambda_\theta \cdot \frac{B}{2} \right) \cdot \dot{q} = \dot{p} \cdot \frac{h}{2} + \Lambda_\theta \cdot \frac{B}{2} \cdot \dot{F}(t), \quad (1)$$

where q is a generalized coordinate corresponding to the angular displacement of the railway ferry with wagons around the longitudinal axis. The origin of the coordinate system is located at the center of gravity of the railway ferry; D is the weight displacement, B is the width of the railway ferry, h is the height of the side of the railway ferry; Λ_θ is the coefficient of vibration resistance, z_g is the

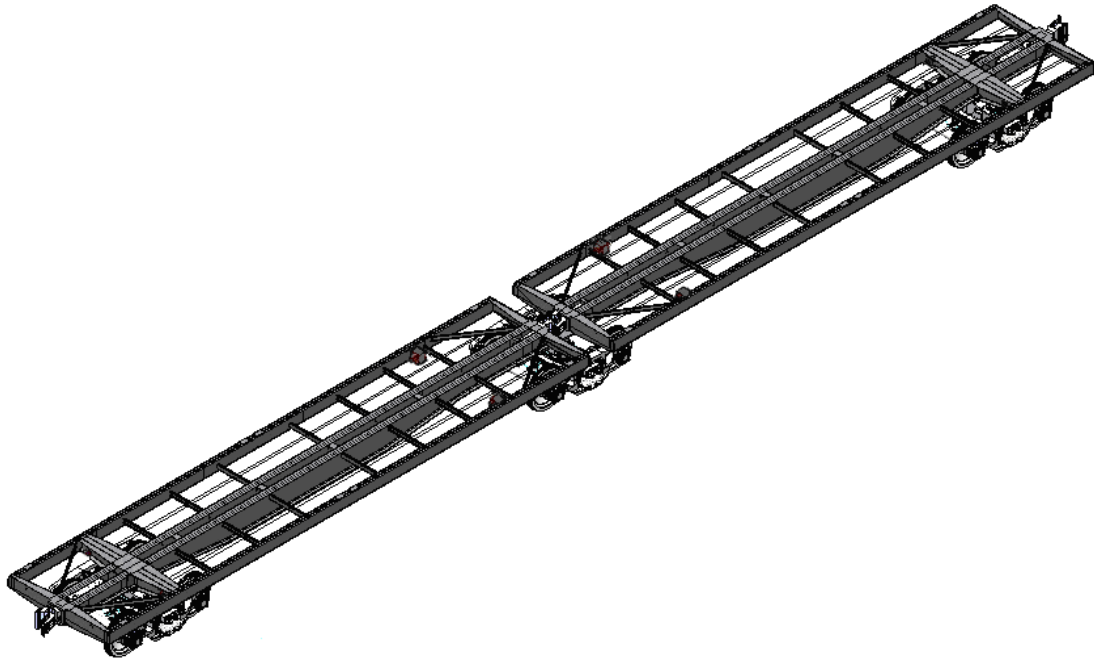


Figure 2. An improved design of the sectional flat wagon

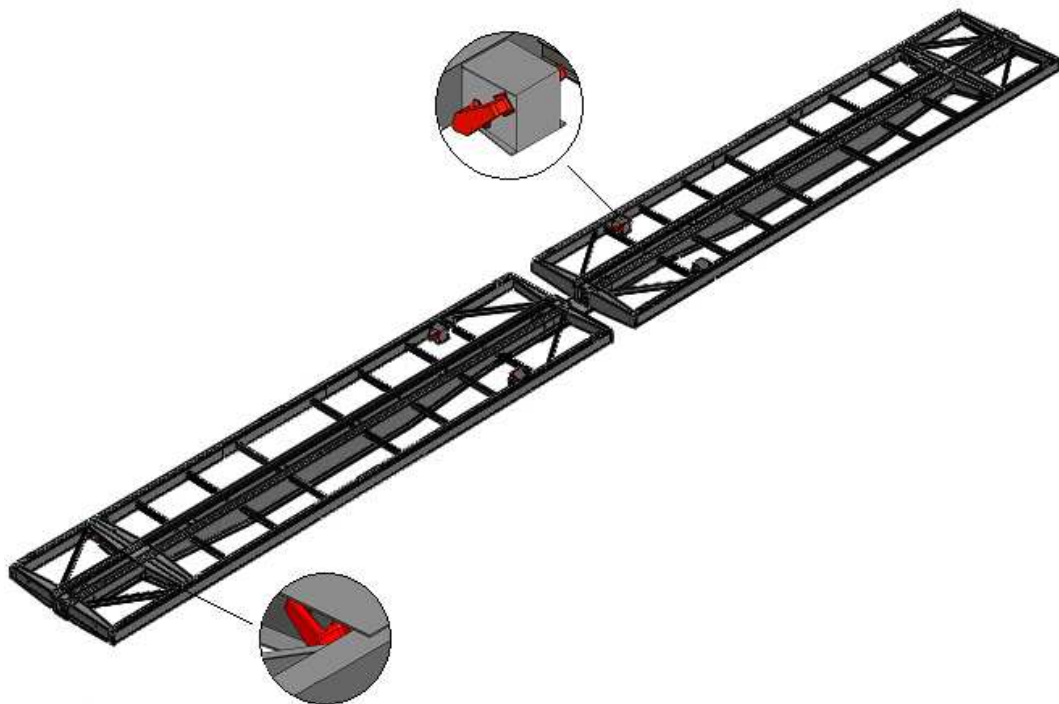


Figure 3. A main load-bearing structure (a frame) of the solved sectional flat wagon

coordinate of the center of gravity of the railway ferry, p' is the wind load, and $F(t)$ is the law of action of the force that disturbs the movement of the train ferry with the wagon bodies placed on its decks.

It was considered for the derived mathematical model that the wagon body is rigidly fixed relative to the reference plane (deck) and moves together

with it. The impact of sea waves on the railway ferry body side, where the wagons are placed on, is not considered.

The model takes into account the wave angles χ relative to the ferry, which are considered in the range of 0° to 180° . These angles were embedded in the frequency of the sea wave, which is considered

in the law of action of the disturbing force, which is described in the form of a trochoid (see Eq.(2)):

$$F(t) = a + R \cdot e^{k \cdot b} \cdot \sin(k \cdot a + \omega \cdot t) + b - R \cdot e^{k \cdot b} \cdot \cos(k \cdot a + \omega \cdot t) \quad (2)$$

where a and b are the coordinates of the center of the principle's trajectory in the horizontal and vertical directions, respectively, R is the radius of the principle's trajectory; ω is the wave frequency, k is the wave trajectory frequency.

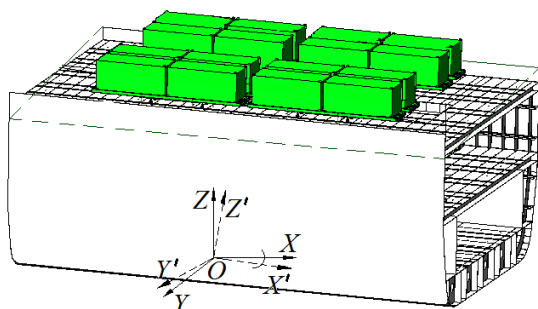


Figure 4. Schemes of movements of railway ferry fragments during roll movement

Input parameters to the derived model (1) are the technical parameters of the considered railway ferry [27], parameters of the sea area [28–30], coordinates of the location of the wagons' bodies relative to the center of oscillations of the ferry deck.

III. RESULTS AND DISCUSSION

A calculation program was developed in the Mathcad package [31–33] to solve the derived differential equation of motion (1). These equations were reduced for this software to the normal Cauchy form. Subsequently, they were integrated using the Runge-Kutta method. The achieved results are shown in **Fig. 5**.

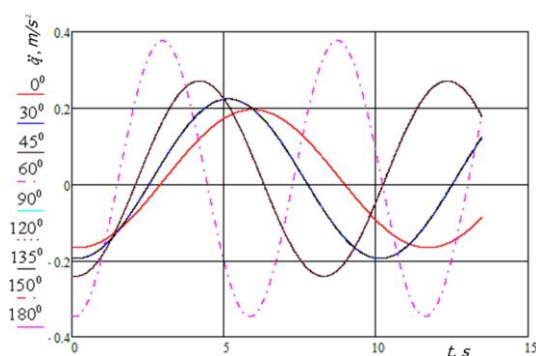


Figure 5. Accelerations acting on the main load-bearing structure of a sectional flat wagon during oscillations of a railway ferry

In this case, the duration of the oscillatory process was taken equal to the wave period, the nature of the disturbance was trochoidal, the amplitude was equal

to the maximum wave height for the given sea area, and the frequency was determined by the course angle of the wave in relation to the railway ferry hull.

Fig. 5 shows the acceleration of platform wagons placed on the track of the railway ferry upper deck furthest from the bulwark during angular movements around the longitudinal axis. The most significant acceleration of the wagons occurs at a course angle of the wave in relation to the railway ferry hull of 0.40 m/s^2 .

The magnitude of the accelerations is given without the g , i.e. without gravitational acceleration.

The results were used in calculations to determine the strength of the main load-bearing structure of the frame of the sectional flat wagon. The calculation was carried out using the finite element method in the CosmosWorks software environment.

A calculation scheme takes into account the vertical loads in the areas of container support on the fitting stops, as well as forces from the chain ties on the fastening nodes relative to the deck (**Fig. 6**). The component forces, which arise from the chain ties and which are determined by the angles of their placement in a space, are listed in **Table 1**.

Table 1. Forces acting on the main load-bearing structure of the frame of a sectional flat wagon through chain ties

Acting force [kN]		
F_x	F_y	F_z
92.94	129.96	160.98

Rigid connections were installed in areas, where the frame rests on mounts.

A graph-analytical method was used to determine the optimal number of grid elements. The number of grid nodes was 148,723, and the number of elements of the finite element mesh was 462,131 with the largest size being of 200 mm and the smallest being of 40 mm. Such element sizes are determined by variational calculations. In this case, the dependence of stresses on the number of finite elements was constructed. When this dependence took the form of a horizontal straight line, it meant that this number of elements is sufficient to obtain reliable calculation results. The results of the calculation of the main load-bearing structure of the flat wagon are presented shown in **Fig. 7**.

The maximum equivalent stresses in the supporting structure of the flat wagon are concentrated in the zone of interaction of the pivot beam with the main longitudinal beam. The maximal value of the stress was 215.4 MPa. However, this value does not exceed the permissible value for the grade of steel of the metal structure [34].

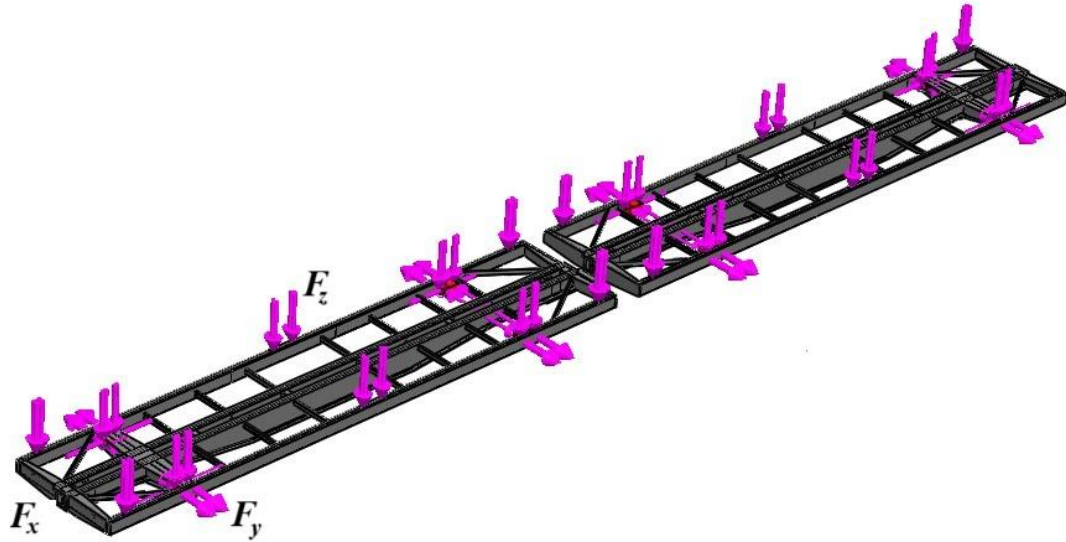


Figure 6. A calculation diagram of the solved sectional flat wagon frame

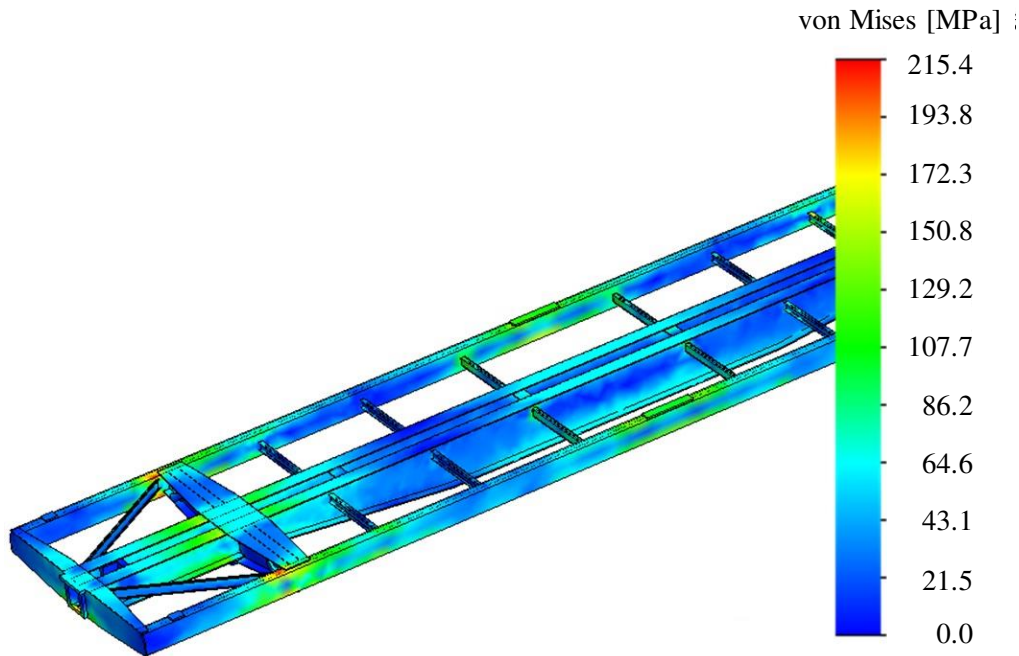


Figure 7. A stress distribution in the solved sectional flat wagon frame

The permissible values of stresses are considered of 310.5 MPa. There is an analogue of this normative document [35]. However, it contains other values permissible stresses. As the study was conducted on the example of a 1520 mm gauge wagon, the document [34] was considered as the basic source.

The further research in this area will be devoted to determining the loading of a flat wagon under other calculation schemes, i.e. rolling onto the ferry (rolling off the ferry), loading during vertical and keel roll of the ferry, the research on the stability of the wagon on the deck, etc. The authors also plan to investigate the fatigue strength of the frame of the flat wagon during transportation by a ferry. A mandatory stage of these studies is also the study of

the loading of 1435 mm gauge wagons during maritime transport.

IV. CONCLUSIONS

1. It is proposed that nodes be installed on their supporting structures for securing chain ties to increase the reliability of securing flat wagons on railway ferries.

2. A mathematical model was developed to determine the numerical values of accelerations as components of the dynamic load acting on the main load-bearing structure of a sectional flat wagon during transportation on a railway ferry. It was established that in the absence of the flat wagon's

movements relative to the deck, the acceleration acting on its supporting structure during angular movements around the longitudinal axis is about 0.40 m/s^2 .

3. The magnitudes of the forces acting on the solved sectional flat wagon frame through the means of securing relative to the decks were determined.

4. The strength of the frame of the sectional flat wagon type during transportation by a rail ferry has been calculated. The maximum equivalent stresses distributed in the frame of the solved wagon do not exceed the permissible ones, and the maximal calculated values are 215 MPa.

5. The conducted research will contribute to improving the safety of maritime transport of wagons.

ACKNOWLEDGEMENT

This research was supported by the projects: VEGA 1/0308/24 "Research of dynamic properties of rail vehicles mechanical systems with flexible components when running on a track" and KEGA 031ŽU-4/2023 "Development of key competencies of the graduate of the study program Vehicles and Engines".

"Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V01-00131."

AUTHOR CONTRIBUTIONS

A. Lovska: Software, Formal Analysis, Investigation, Data curation, Visualisation, Project Administration.

J. Gerlici: Conceptualization, Methodology, Formal analysis, investigation, Writing – original draft preparation.

J. Dižo: Methodology, Validation, Investigation, Data curation, Writing – review and editing, Project Administration.

V. Ravlyuk: Software, Formal Analysis, Investigation, Resources, Data curation, Visualisation.

A. Rybin: Validation, Investigation, Supervision, Writing – review and editing, Visualisation, Project administration.

DISCLOSURE STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID

A. Lovska <http://orcid.org/0000-0002-8604-1764>

J. Gerlici <https://orcid.org/0000-0003-3928-0567>

J. Dižo <http://orcid.org/0000-0001-9433-392X>

V. Ravlyuk <https://orcid.org/0000-0003-4818-9482>

A. Rybin <https://orcid.org/0000-0001-7546-0077>

REFERENCES

- [1] S. Panchenko, J. Gerlici, G. Vatulia, A. Lovska, A. Rybin, O. Kravchenko, Strength assessment of an improved design of a tank container under operating conditions, Communication-Scientific Letters of the University of Zilina 25 (3) (2023) pp. B.186–B.193. <https://doi.org/10.26552/com.C.2023.047>
- [2] G. Vatulia, A. Lovska, M. Pavliuchenkov, V. Nerubatskyi, A. Okorokov, D. Hordiienko, R. Vernigora, I. Zhuravel, Determining patterns of vertical load on the prototype of a removable module for long-size cargoes, Eastern-European Journal of Enterprise Technologies 120 (6/7) (2022) pp. 21-29. <https://doi.org/10.15587/1729-4061.2022.266855>
- [3] G. Divya Priya, Modeling and analysis of twenty tonne heavy duty trolley, International Journal of Innovative Technology and Research 2 (6) (2014) pp. 1568-1580. <http://www.ijitr.com/index.php/ojs/article/view/421/pdf>
- [4] W. Krason, T. Niezgoda, Fe numerical tests of railway wagon for intermodal transport according to PN-EU standards, Bulletin of the Polish Academy of Sciences Technical Sciences 62 (4) (2014) pp. 843-851. <https://doi.org/10.2478/bpasts-2014-0093>
- [5] S. Myamlin, A. V. Shatunov, A. V. Sorokolet, Rolling stock for the transportation of containers by rail, Collection of scientific works of Donetsk Institute of Railway Engineering (in Russian), 2010, pp. 125–132.
- [6] WBN Waggonbau Niesky GmbH: Developing a flexible platform of freight wagons, Intern. Edition, 1, 2016, p. 46.
- [7] K. Wójcik, J. Malachowski, P. Baranowski, Ł. Mazurkiewicz, K. Damaziak, W. Krasoń, Multi-body Simulations of Railway Wagon Dynamics, Journal of KONES Powertrain and

- Transport 19 (3) (2015), pp. 499-506.
<https://doi.org/10.5604/12314005.1138164>.
- [8] N. L. Gurzhy, Improving the technical characteristics of a sectional flat wagon by improving the design (in Ukrainian), The author's abstract of the dissertation candidate of technical sciences, Dnipropetrovsk National University of Railway Transport named after Acad. V. Lazaryan, Dnipropetrovsk, 2010, p. 20.
- [9] A. Das, G. Agarwal, Investigation of Torsional Stability and Camber Test on a Meter Gauge Flat Wagon, *Advances in Fluid Mechanics and Solid Mechanics* (2020), pp. 271–280.
https://doi.org/10.1007/978-981-15-0772-4_24
- [10] L. Ézsiás, R. Tompa, S. Fischer, Investigation of the possible correlations between specific characteristics of crushed stone aggregates, *Spectrum of Mechanical Engineering and Operational Research 1* (1) (2024), pp. 10–26.
<https://doi.org/10.31181/smeor1120242>.
- [11] S. Fischer, Investigation of the settlement behavior of ballasted railway tracks due to dynamic loading, *Spectrum of Mechanical Engineering and Operational Research 2* (1) (2024), pp. 24–46.
<https://doi.org/10.31181/smeor21202528>.
- [12] S. Fischer, D. Harangozó, D. Németh, B. Kocsis, M. Sysyn, D. Kurhan, A. Brautigam, Investigation of heat-affected zones of thermite rail welding, *Facta Universitatis, Series: Mechanical Engineering 22* (4) (2024), pp. 689–710.
<https://doi.org/10.22190/FUME221217008F>
- [13] S. Szalai, B. F. Szívós, D. Kocsis, M. Sysyn, J. Liu, and S. Fischer, The application of DIC in criminology analysis procedures to measure skin deformation, *Journal of Applied and Computational Mechanics*, 10 (4) (2024), pp. 817–829.
<https://doi.org/10.22055/jacm.2024.46966.4634>.
- [14] S. Fischer, S. Kocsis Szürke, Detection process of energy loss in electric railway vehicles, *Facta Universitatis, Series: Mechanical Engineering 21* (1) (2023), pp. 81–99.
<https://doi.org/10.22190/FUME221104046F>.
- [15] S. Kocsis Szürke, G. Kovács, M. Sysyn, J. Liu, S. Fischer, Numerical optimization of battery heat management of electric vehicles, *Journal of Applied and Computational Mechanics 9* (4) (2023) pp. 1076–1092.
<https://doi.org/10.22055/jacm.2023.43703.4119>.
- [16] S. Abdullayev, G. Bakyt, A. Kamzina, K. Sarsanbekov, A. Abdullayeva, Interaction of the TE33a Diesel Locomotive and the Railway Track on Curved Section with Radius 290 m, *Communications - Scientific Letters of the University of Žilina 25* (4) (2023), pp. B315–B326.
<https://doi.org/10.26552/com.C.2023.069>
- [17] E. Szczepański, P. Gołębiowski, B. Kondracka, Evaluation of the technological process of wagon processing at shunting stations using the simulation model, *Scientific Journal of Silesian University of Technology. Series Transport*, 120 (2023), pp. 249–267.
<https://doi.org/10.20858/sjsutst.2023.120.16>.
- [18] P. Fabian, J. Gerlici, J. Masek, P. Marton, Versatile, efficient and long wagon for intermodal transport in Europe, *Communications-Scientific Letters of the University of Žilina 15* (2) (2013) pp. 118–123.
<https://doi.org/10.26552/com.c.2013.2.118-123>
- [19] M. Kostrzewski, A. Kostrzewski, Analysis of operations upon entry into intermodal freight terminals, *Applied Sciences 9* (121) (2019) 558.
<https://doi.org/10.3390/app9122558>
- [20] J. Gnap, Š. Senko, M. Kostrzewski, M. Brídžiková, R. Czödöröová, Z. Říha, Research on the relationship between transport infrastructure and performance in rail and road freight transport—a case study of Japan and selected European countries, *Sustainability 13* (122) (2021) 6654.
<https://doi.org/10.3390/su13126654>
- [21] P. Šťastniak, L. Smertanka, M. Moravčík, Structural analysis of a main construction assemblies of the new wagon prototype type Zans, *Manufacturing Technology 18* (3) (2018), pp. 510–517.
<https://doi.org/10.21062/ujep/130.2018/a/1213-2489/MT/18/3/510>
- [22] M. Opala, J. Korzeb, S. Koziak, R. Melnik, Evaluation of stress and fatigue of a rail vehicle suspension component, *Energies 14* (12) (2021) 3410.
<https://doi.org/10.3390/en14123410>
- [23] K. Nehring, M. Lasota, A. Zabielska, R. Jachimowski, A multifaceted approach to assessing intermodal transport, *Scientific Journal of Silesian University of Technology. Series Transport 121* (2023), pp. 141–165.
<https://doi.org/10.20858/sjsutst.2023.121.10>.
- [24] R. Melnik, B. Sowinski, Application of the rail Vehicle's monitoring system in the process of suspension condition assessment, *Communications - Scientific Letters of the University of Žilina 15* (4) (2013) pp. 3–8.
<https://doi.org/10.26552/com.c.2013.4.3-8>
- [25] W. Nawrocki, R. Stryjski, M. Kostrzewski, W. Woźniak, T. Jachowicz, Application of the vibro-acoustic signal to evaluate wear in the spindle bearings of machining centres. In-service diagnostics in the automotive industry, *Journal of Manufacturing Processes 92* (2023), pp. 165–178.

- <https://doi.org/10.1016/j.jmapro.2023.02.036>
- [26] A. O. Lovska, Articulated flat wagon for container transportation, Ukrainian patent for an invention, No. 22328, MPK (2006.01) B61D 3/08 (2006.01), B61D 3/10 (2006.01) B61D 3/20 (2006.01) B60P 7/13 (2006.01) B60P 7/08 (2006.01) B61F 1/08 (2006.01) B61F 1/02 (2006.01), a201704241; application 28.04.2017; publ. 26.10.2020, Bull. No. 20.
- [27] Cargo securing manual for m/v "Geroy Shipky" (in Ukrainian) No. 2512. 02. – Official ed. – Odessa: Ministry of Transport of Ukraine. State Department of Maritime and River Transport. 1997. 51 p.
- [28] Wind and waves in the oceans and seas: reference data [ed. by I.N. Davidan] – L. (in Ukrainian): Transport (1974), 360 p.
- [29] E. Di Gialleonardo, S. Melzi, D. Trevisi, Freight trains for intermodal transportation: optimisation of payload distribution for reducing longitudinal coupling forces, *Vehicle System Dynamics* 61 (10) (2023), pp 2532–2550.
<https://doi.org/10.1080/00423114.2022.2120025>
- [30] J. Caban, F. Brumerčik, J. Vrábel, P. Ignaciuk, W. Misztal, A. Marczuk, Safety of maritime transport in the Baltic Sea, *MATEC Web of Conferences* 134 (2017) 13400003.
<https://doi.org/10.1051/mateconf/201713400003>
- [31] I. V. Bogach, O. Yu. Krakovetskyi, L. V. Kylyk, Numerical methods for solving differential equations using MathCad: Textbook (in Ukrainian), Vinnytsia, VNTU, (2020), 106 p.
- [32] A. V. Syasev, Introduction to the MathCad system: a textbook (in Ukrainian), Dnipropetrovsk: Publishing house of Dnipropetrovsk University (2004), 108 p.
- [33] J. Gerlici, A. Lovska, G. Vatulia, M. Pavliuchenkov, O. Kravchenko, S. Solcansky, Situational adaptation of the open wagon body to container transportation, *Applied Sciences* 13 (15) (2023), 8605.
<https://doi.org/10.3390/app13158605>
- [34] DSTU 7598:2014. Freight wagons. General requirements for calculations and design of new and modernized 1520 mm gauge wagons (non-self-propelled). Kyiv, 2015. 162 p.
- [35] EN 12663-2. Railway applications – structural requirements of railway vehicle bodies – Part 2: Freight wagons. Bulgaria, 2010. 54 p.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license.