

Estimation of deformation modulus of coals using artificial neural networks (ANN)

Ekin Köken^{1,*}

¹Nanotechnology Engineering Department, Abdullah Gul University
Address, 38100, Kayseri, Turkey
*e-mail: ekin.koken@agu.edu.tr

Submitted: 23/04/2022 Accepted: 19/05/2022 Published online: 29/05/2022

Abstract: In this study, the Young modulus (E) of different coals was investigated using artificial neural networks (ANN). For this purpose, a comprehensive literature survey was carried out to compile such datasets available for the ANN analyses. As a result of the literature survey, a database composed of 81 datasets was formed. In the ANN analyses, uniaxial compressive strength (UCS) and dry density (ρ_d) of coals were adopted as input parameters. The ANN analysis results demonstrated that the predictive model established in this study could be reliably used to estimate the E for different coals. The correlation of determination value (R^2) for the developed model is 0.85, which shows its relative success. In this context, this study can be declared a case study showing the applicability of ANN for the evaluation of E for a wide range of coal types. However, the number of samples and independent variables should be increased to obtain more comprehensive models in future studies.

Keywords: Coal; Deformation properties; Young modulus; Artificial neural networks

I. INTRODUCTION

The stability of coal measures strata in underground mines is of prime importance in sustainable and safe coal mining operations. The stability of coal measures strata has been investigated mainly using numerical modeling techniques [1–3]. In numerical modeling of rock masses, several rock strength properties such as uniaxial compressive strength (UCS), Young modulus (E), and Poisson's ratio (ν) are required as input parameters. By adopting the above-mentioned rock properties, coal-bearing rock masses can be modeled using several methodologies such as the finite element method (FEM) and the discrete element method (DEM) [4–6]. Of the rock properties, the UCS and E are of prime importance to set forth the stability of coal pillars and the stress-strain relationship of coal-bearing strata. However, considering the heterogeneous structure of coals, the determination of E in the laboratory is tedious and requires special equipment such as high precision stiff loading machines, strain gauges, deformation jackets, or linear variable differential transformers (LVDTs). Hence, several theories have been postulated to estimate the E of different rock types in the literature [7–11]. Recently, soft computing algorithms such as adaptive neuro-fuzzy inference

systems (ANFIS) and artificial neural networks (ANN) have gained popularity in dealing with most engineering geological problems because of their flexibility and high precision accuracy. [12–16].

However, apart from the studies by Pan et al. [17] and Lawal et al. [18], the implementation of regression and soft computing tools for the evaluation of E for different coals is quite limited. Therefore, there is a need to obtain comprehensive empirical models to evaluate the E of different coal types. For this purpose, a comprehensive literature survey was conducted to compile such datasets for soft computing analyses in this study. The most important theoretical and practical findings obtained from this literature survey are summarized as follows:

- The strength properties of coals increase in parallel with their rank [17].
- The variations in pulse wave velocity (V_p) are highly dependent upon the E of coals [19].
- The UCS of coals can be estimated from Schmidt Hammer tests [20–23].

However, the above findings are valid mainly for a small area of interest. Therefore, they have some limitations in dealing with larger datasets with different coal origins. Consequently, soft computing

analyses with larger datasets are required to obtain more comprehensive relationships for the evaluation of coal strength and deformation properties.

In this study, the E of different coals was investigated using ANN analyses. On the basis of the collected data, a comprehensive empirical model is introduced. The details and mathematical expressions of the established model are also presented in this study to allow users to implement the proposed model more efficiently.

II. DATABASE DEVELOPMENT

A comprehensive literature survey was conducted to compile quantitative data on the strength and deformation properties of different coal types. Unfortunately, a significant number of previous studies could not have been considered due to a lack of information on the physical and mechanical properties of coal, which are important as input parameters. As a result of the literature survey, a database composed of 81 datasets was collected including the dry density (ρ_d), UCS, and E (**Table 1**).

Table 1. Datasets considered in this study

ρ_d [g/cm ³]	UCS [MPa]	E [GPa]	n	Ref.
1.75–2.15	8.207–54.702	1.457–3.213	4	[19]
1.40–1.90	31.01–33.00	3.52–3.70	2	[24]
1.83–1.89	34.12–35.68	2.38–2.41	6	[25]
1.20–1.70	6.75–22.30	0.23–0.78	9	[26]
1.27–1.80	3.08–28.77	0.81–3.82	15	[27]
1.37–1.98	17.38–32.39	2.19–2.43	45	[28]

Using the database summarized in Table 1, several ANN analyses were performed to establish a comprehensive mathematical model for the evaluation of the E of different coals. Before performing the ANN analyses, the simple correlations of the considered variables were revealed by Pearson’s correlation coefficient (r) and spearman rho values, which are listed in **Table 2**. Accordingly, the ρ_d and UCS are moderately associated with the E of different coal types. Therefore, these two independent variables were selected as input parameters in ANN analyses.

Table 2. Correlations of independent variables for the evaluation of E for different coal types

Statistical indicator	ρ_d	UCS
Pearson’s correlation coefficient, r	0.566	0.565
Spearman rho value	0.766	0.500

III. ARTIFICIAL NEURAL NETWORKS (ANN) ANALYSES

ANN-based methods can analyze data, learn, save knowledge, and use it for future predictions [29, 30]. This parallel distribution learning algorithm applies to many problems, from social science to applied science. In most ANN models, a feedforward backpropagation algorithm is adopted. In this study, the neural network toolbox (nntool) was utilized to establish several neural networks in the MATLAB environment. Various possible network architectures with variable hidden layers and neurons were attempted to determine the most reliable ANN structure. For estimating the E for different coal types, the most convenient ANN architecture was found to be 2–6–1 (**Fig. 1**). The independent variables for the ANN analyses were selected as the UCS and ρ_d . To increase training efficiency during ANN analyses, the dataset was also normalized between –1 and 1, using equation (1).

$$V_n = 2 \cdot \frac{x_i - x_{min}}{x_{max} - x_{min}} - 1 \quad (1)$$

where x_i is the relevant parameter to be normalized, x_{min} , and x_{max} are the minimum and maximum values in the dataset.

As a result of the ANN analyses, the E for different coal types can be estimated using equation (2). The subfunctions of equation (2) were determined based

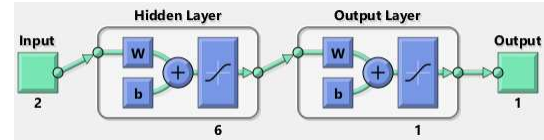


Figure 1. ANN architecture adopted in this study

on the deterministic approach previously described by Das [31] and they are listed in equations (3) to (10), where equations (9) to (10) are the normalization functions. According to the ANN analyses, the proposed model (equation (2)) correlates with a determination value (R^2) of 0.85, which shows its relative success.

$$E = 1.8233 \cdot \tanh\left(\sum_{i=1}^6 A_i + 1.1645\right) + 2.0264; R^2 = 0.85 \quad (2)$$

$$A_1 = 1.0654 \cdot \tanh(7.5356^n \cdot UCS - 1.6385^n \cdot \rho_d - 1.6398) \quad (3)$$

$$A_2 = 0.85969 \cdot \tanh(-14.0079^n \cdot UCS + 7.484^n \cdot \rho_d - 0.5834) \quad (4)$$

$$A_3 = -4.787 \cdot \tanh(1.6683^n \cdot UCS - 4.2348^n \cdot \rho_d - 2.1176) \quad (5)$$

$$A_4 = 2.4436 \cdot \tanh(0.56564^n \cdot UCS + 3.5425^n \cdot \rho_d - 0.43686) \quad (6)$$

$$A_5 = -7.7371 \cdot \tanh(-1.2373^n \cdot UCS + 2.4793^n \cdot \rho_d - 0.93343) \quad (7)$$

$$A_6 = 0.25474 \cdot \tanh(4.5541^n \cdot UCS - 4.5541^n \cdot \rho_d + 8.0908) \quad (8)$$

$${}^n UCS = 0.0387 \cdot UCS - 1.1193 \quad (9)$$

$${}^n \rho_d = 2.1075 \cdot \rho_d - 3.5311 \quad (10)$$

IV. RESULTS AND DISCUSSION

Based on the ANN analyses, the proposed empirical model (Eq 2) was developed in this study. The performance of the proposed model was checked by correlating the predicted and measured E values for different coal types, which were previously reported by several researchers.

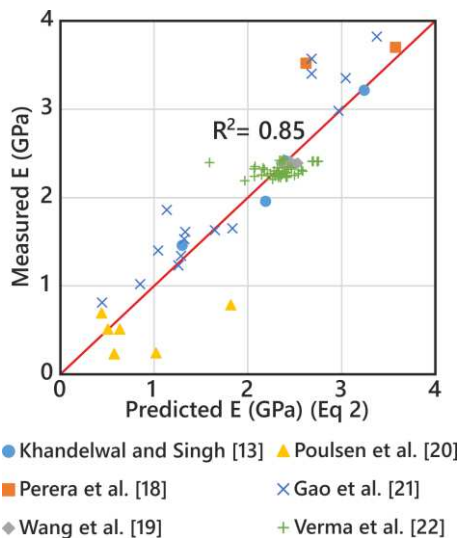


Figure 2. Predicted and measured E values for the proposed ANN model

Consequently, the predicted and measured E values for different types of coal are in good agreement (Fig. 2). Therefore, the model established in this study can be reliably used to estimate the E of different coals.

The deformation properties of coals are well-known phenomena for coalbed methane recovery and CO₂ sequestration [32]. They are also important in estimating the bearing capacity of the coal masses [23]. Therefore, comprehensive models are needed for the evaluation of E for different coal types. In most engineering projects related to the underground coal mines, the UCS and ρ_d values have been measured routinely. Hence, the E of different coal

types can also be estimated using these coal properties. In this context, this study can be declared a case study showing the applicability of ANN for the evaluation of E for a wide range of coal types. However, the number of samples and independent variables should be increased to obtain more comprehensive models in future studies. Anyhow, the present study can be declared a case study showing the applicability of ANN analyses for the evaluation of E for different coal types.

V. RESULTS AND DISCUSSION

The E of coals is a fundamental parameter for determining the deformation behavior of the coal masses. However, because of the heterogeneity and complexity of the coal strata, the determination of E for different types of coal is challenging and requires special equipment in the laboratory. Therefore, it is necessary to obtain reliable and comprehensive models to estimate E for a wide range of rock types. With this study, a comprehensive predictive model is introduced to estimate the E of different coals. For this purpose, a comprehensive literature survey was carried out to compile datasets available for soft computing analyses. Consequently, a database composed of 81 datasets was formed (Table 1). Soft computing analyzes based on ANN were then performed to build a novel predictive model for the evaluation of E for different coals.

In the ANN analyses, the UCS and ρ_d values of coals were considered as input parameters. As a result of ANN analyzes, Eq. 2 was developed, which successfully estimated the E of the coals. For the sake of clarity, the sub equation systems behind equation (2) were also presented in this paper (equations (3) to (10)), to let users implement the proposed model efficiently. According to the performance of the proposed ANN model, it was determined that the predicted and measured E values are in good agreement (Fig. 2), indicating the relative success of the model. However, the number of samples and independent variables should be increased to obtain more comprehensive models in future studies.

AUTHOR CONTRIBUTIONS

E. Köken: Conceptualization, Software, Writing, Review, Validation and Editing.

DISCLOSURE STATEMENT

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this document.

ORCID

E. Köken <http://orcid.org/0000-0003-0178-329X>

REFERENCES

- [1] X. Liu, X. Zhang, H. Wang, B. Jiang, Laboratory testing and analysis of dynamic and static resilient modulus of subgrade soil under various influencing factors, *Construction and Building Materials* 195 (2019) pp. 178–186. <https://doi.org/10.1016/j.conbuildmat.2018.11.061>
- [2] X. Chu, A Review on the Resilient Response of Unsaturated Subgrade Soils, *Advances in Civil Engineering* (2020), 7367484 <https://doi.org/10.1155/2020/7367484>
- [3] F. Lekarp, U. Isacsson, and A. Dawson, "State of the Art. I: Resilient Response of Unbound Aggregates," *Journal of Transportation Engineering-asce - J TRANSP ENG-ASCE* 126 (1) (2000) [https://doi.org/10.1061/\(ASCE\)0733-947X\(2000\)126:1\(66\)](https://doi.org/10.1061/(ASCE)0733-947X(2000)126:1(66))
- [4] A. S. El-Ashwah, A. M. Awed, S. M. El-Badawy, and A. R. Gabr, "A new approach for developing resilient modulus master surface to characterize granular pavement materials and subgrade soils," *Construction and Building Materials* 94 (2019) pp. 372-385.
- [5] Y. Yao, J. Zheng, J. Zhang, J. Peng, and J. Li, "Model for predicting resilient modulus of unsaturated subgrade soils in south China," *KSCE Journal of Civil Engineering* 22 (6) (2018) pp. 2089-2098.
- [6] K. Naji, "Resilient modulus–moisture content relationships for pavement engineering applications," *Int. J. Pavement Eng.* 19 (7) (2018) pp. 651-660.
- [7] Y. J. Cui, "Mechanical behaviour of coarse grains/fines mixture under monotonic and cyclic loadings," *Transportation Geotechnics, Review* 17 (2018) pp. 91-97. <https://doi.org/10.1016/j.trgeo.2018.09.016>
- [8] J. Zhang, J. Peng, L. Zeng, J. Li, and F. Li, "Rapid estimation of resilient modulus of subgrade soils using performance-related soil properties," *Int. J. Pavement Eng.* (2019) pp. 1-8.
- [9] C. E. Cary and C. E. Zapata, "Resilient modulus for unsaturated unbound materials," *Road Materials and Pavement Design* 12 (3) (2011) pp. 615-638.
- [10] H. Park, G. Kweon, and S. R. Lee, "Prediction of resilient modulus of granular subgrade soils and subbase materials using artificial neural network," *Road Materials and Pavement Design* 10 (3) (2009) pp. 647-665.
- [11] S. Jayakody, C. Gallage, and J. Ramanujam, "Performance characteristics of recycled concrete aggregate as an unbound pavement material," *Heliyon, Review* 5 (9) (2019), Art no. e02494 <https://doi.org/10.1016/j.heliyon.2019.e02494>
- [12] N. Su, F. Xiao, J. Wang, and S. Amirkhanian, "Characterizations of base and subbase layers for Mechanistic-Empirical Pavement Design," *Construction and Building Materials, Review* 152 (2017) pp. 731-745. <https://doi.org/10.1016/j.conbuildmat.2017.07.060>
- [13] J. H. Zhang, J. H. Peng, and J. L. Zheng, "Progress and Prospect of the Prediction Model of the Resilient Modulus of Subgrade Soils," *Zhongguo Gonglu Xuebao/China Journal of Highway and Transport, Review* 33 (1) (2020) pp. 1-13. <https://doi.org/10.19721/j.cnki.1001-7372.2020.01.001>
- [14] J. M. Rasul, G. S. Ghataora, and M. P. Burrow, "The effect of wetting and drying on the performance of stabilized subgrade soils," *Transportation Geotechnics* 14 (2018) pp. 1-7.
- [15] X. R. Wu and L. L. Zhu, "Analysis of the influence of water content on shanxi loess CBR and rebound module," in H. Abbas and T. K. Hwee Eds., *Advanced Materials Research. Durnten-Zurich: Trans Tech Publications Ltd* 919-921 (4) (2014), pp. 820-823.
- [16] R. Ji, N. Siddiki, T. Nantung, and D. Kim, "Evaluation of resilient modulus of subgrade and base materials in Indiana and its implementation in MEPDG," *The Scientific World Journal* 2014 (2014).
- [17] F. Salour and S. Erlingsson, "Resilient modulus modelling of unsaturated subgrade soils: laboratory investigation of silty sand subgrade," *Road Materials and Pavement Design* 16 (3) (2015) pp. 553-568.
- [18] X. R. Wu and L. L. Zhu, "Study on the Relations of Strength control indicators of road subgrade in shanxi loess region," in H. Abbas and T. K. Hwee Eds., *Advanced Materials Research. Durnten-Zurich: Trans Tech Publications Ltd* 919-921 (2014) pp. 1160-1163.
- [19] F. Salour, S. Erlingsson, and C. E. Zapata, "Modelling resilient modulus seasonal variation of silty sand subgrade soils with matric suction control," *Can. Geotech. J.* 51 (12) (2014) pp. 1413-1422.
- [20] F. Achampong, M. Usman, and T. Kagawa. Evaluation of resilient modulus for lime- and cement-stabilized synthetic cohesive soils, *Transportation Research Record* (1997) pp. 70-75.
- [21] A. Rahim and K. George, "Models to estimate subgrade resilient modulus for pavement design," *Int. J. Pavement Eng.* 6 (2) (2005) pp. 89-96.
- [22] C. N. Khoury, N. N. Khoury, and G. A. Miller. Effect of cyclic suction history (hydraulic hysteresis) on resilient modulus of unsaturated

- fine-grained soil, Transportation Research Record (2011) pp. 68-75.
- [23] AASHTO, "AASHTO M 145 (1991), AASHTO Standard Specification for Classification of Soils and Soil–Aggregate Mixtures for Highway Construction Purposes," 1991.
- [24] A. C. D.-o. Soil and Rock, Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System) 1. ASTM international, 2017.
- [25] AASHTO, "AASHTO T 307-99 (2007), AASHTO Standard method of test for determining the resilient modulus of soils and aggregate materials, Washington D.C.," 2009.
- [26] T. R. Board, E. National Academies of Sciences, and Medicine, Laboratory Determination of Resilient Modulus for Flexible Pavement Design. Washington, DC: The National Academies Press (in English), 2004.
- [27] AASHTO, "AASHTO T 274 (1994), AASHTO Standard Method of Test for Resilient Modulus of Unbound Granular Base/Subbase Materials and Subgrade Soils - SHRP Protocol P46," 1994.
- [28] ASTM, "ASTM D 559-96, (2003), Standard test method for wetting and drying of compacted mixture, ASTM International, Pennsylvania, USA."



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.

A comparative analysis of road and rail freight transport through the Republic of Serbia from the aspect of external costs

Nikola Petrović*, Vesna Jovanović, Boban Nikolić, Jovan Pavlović, Jelena Mihajlović

Faculty of Mechanical Engineering, University of Niš
Aleksandra Medvedeva 14, 18106, Niš, Serbia
*e-mail: petrovic.nikola@masfak.ni.ac.rs

Submitted: 29/03/2022 Accepted: 22/05/2022 Published online: 30/05/2022

Abstract: Transport causes significant external effects that have a very harmful impact on the environment, human health, but also the economy. The costs of these external effects are not borne directly by those who caused them, but by other road users (congestion, accidents) and society as a whole (environmental pollution costs). This results in making wrong decisions in the transport market, which further leads to an inefficient use of available resources and loss of social welfare. The purpose of this paper is a comparative analysis of negative external costs arising from freight road and rail transport. The main goal of the paper is to quantify the external costs caused by two different modes of freight transport, based on the characteristics of the transport vehicles and the characteristics of the goods being transported, as well as the length of the goods transport. The road and rail freight transport through Serbia will be chosen as the case study.

Keywords: External costs; Internalisation; Modes of transport

I. INTRODUCTION

Transport significantly contributes to economic growth and market development, but in most cases, some transport modes have not only positive but also very negative effects on society itself [1]. The activity of transport results in negative external effects such as emissions, noise, congestion and infrastructural wear and tear [2]. The costs of these negative effects of transport are generally not borne by transport users, and are therefore not taken into account when deciding on the choice of transport mode. External costs are incurred by social and economic activities, in this case transport, of one group of persons, so that they are reimbursed by another group of persons, which appear as users, direct or indirect, of performed activities, i.e. transport services of the first group of persons.

In order to correctly define external costs, it is first necessary to make a difference between social and private costs. Social costs reflect all costs incurred due to the provision and use of transport infrastructure (costs of wear and tear of infrastructure, capital costs, congestion costs, costs of traffic accidents, etc.). Private costs originate directly from transport users (vehicle maintenance and energy costs, vehicle use costs, transport tariffs, taxes, fees, etc.)

The total external transport costs of the European Union countries by transport modes for 2019 amount to a total of 987 billion euros [3]. The largest part of these costs, 83%, goes to road transport, while other transport modes cost significantly less, 10% for sea, 5% for air, 1.8% for rail and only 0.3% for inland waterway transport. External costs, for 2019, incurred by transport activities were allocated to passenger transport in the amount of €625.2 billion and to freight transport in the amount of €195.1 billion. In addition, external transport costs account for a total of 5.7% of GDP (gross domestic product) in the European Union, i.e. 4.3% of passenger transport and 1.4% of freight transport [3]. The highest external transport costs, expressed in billions of euros, are in Germany, Italy, France, Great Britain and Spain, and the lowest in Slovenia, Latvia, Estonia, Cyprus and Malta.

The issue of external effects of transport in the Republic of Serbia is extremely complex, primarily due to lack of funds, regulatory mechanisms, ethics, environmental culture, etc. Despite the complexity of the problem, there are a number of reasons to solve it. Bearing in mind that external transport costs are difficult to define, quantify, and charge through certain economic mechanisms, the theory and practice in Serbia draw on the European experience which shows that economic, environmental and

transport policies are integrated to solve the problem of external transport costs.

Emissions of pollutants in the Republic of Serbia for the period from 1990 to 2018, in terms of types of pollutants, were decreasing from year to year, although this reduction was initially faster and more intense, while for the period from 2015 to 2018 one could say that the percentage of pollutant emissions was more or less stagnant, and that it was in balance, without a tendency to decline significantly [4].

In the freight transport of the Republic of Serbia, road transport dominates when it comes to external costs in all categories. External costs of road transport caused by traffic accidents and air pollution are many times higher than for other transport modes.

The internalisation of external costs using market-based instruments is considered a very effective way to limit the negative effects of transport. Transport users will certainly alter their behaviour in the form of changes in vehicle type, transport vehicle, transport mode, etc., when they can benefit from it.

The aim of this paper is to quantify and perform a comparative analysis of the external costs of road and rail freight transport from the perspective of air pollution and accidents on a practical example in the Republic of Serbia.

II. LITERATURE REVIEW

Based on the research in scientific databases, various techniques and procedures are used to calculate external costs separately for all transport modes from well-known different negative influences (air pollution, accidents, noise, climate change, etc.).

Transport costs are often analysed by different macro and micro economic models, but detailed cost calculations within transport companies are difficult to publish although this information is necessary to control transport operations. Traditional and improved costing approaches can be implemented in transport management. The principle of multistage allocation of total costs was adjusted to the specific characteristics of the transport used in the practical example in [5]. The model in [6] was developed for predicting the volume of railway transport that can be applied in different economic contexts and used as a means of transport planning.

In paper [7], the authors dealt with the creation and application of the calculator of external costs of transport services as an extension, i.e. emission calculator for transport services. Manufacturing and trading companies require carriers to report the environmental impact of their activities. Areas of concern are emission factors used to calculate emissions and aggregate them. The STN EN 16258

standard was used, which deals with the methodology of calculation and declaration of energy consumption and greenhouse gas emissions from transport services due to the fact that the existing emission calculators do not allow the calculation of external costs of transport services in the field of environmental transport services.

A large number of negative externalities are associated with freight transport. Paper [8] applied a quantitative model that explores and analyses the scenario of a coordinated EU approach to internalise the external transport costs of trade activities outside the EU. The results show a positive effect on real GDP and employment in the EU, provided that revenues from these trade charges are recycled back into the economy. The analysis identified policy options that ensure that transport prices reflect social costs. Several approaches are used to reduce external costs, such as prohibitions, regulations, taxes, levies and exchangeable permits. At the EU level, external costs related to the transport of goods are only partially reflected in transport prices.

In many developing countries, such as Mexico, there are no monetary estimates of the external effects of road transport. In paper [9], six categories of estimates were calculated using available data and well-established methods. The results showed that road externalities amounted to at least 59.42 billion US dollars per year or 6.24% of GDP. Per components, accidents accounted for the largest share (28%), followed by congestion (22%), greenhouse gases (21%), air pollution (13%), infrastructure (7%) and noise (9%).

A simulation-based approach for calculating and internalising correct dynamic price levels using two external factors simultaneously is given in [10]. For a real-world case study, it was shown that an iterative price calculation based on cost estimates from the literature makes it possible to identify the amplitude of the correlation between the two external effects under consideration. Economic assessment indicators for the common price policy make it possible to compare other policies with this reference state of the transport system.

Companies working to reduce their environmental impact operate on three levels: optimising existing networks and flows; transport mode optimisation; increasing the efficiency of routes and travel. Many actions aimed at reducing the cost of transport pollution include minimising empty trucks, encouraging cooperative retail distribution, launching more efficient vehicles; all measures that, before reducing pollution and congestion costs, have a significant benefit of reducing operating costs paid directly by the companies. Paper [11] reviewed the literature and analytical aggregation of different results, in order to obtain a homogeneous function of

transport costs, and then introduced several applications to explain the proposed models.

The use of the multi-criteria method, Analytical Hierarchical Process (AHP), to rank different strategies for reducing emissions on the road, including strategies to reduce, avoid and replace, was carried out in paper [12]. The data were obtained on the basis of a survey of experts in the field of transport and climate sciences. The aim of the model was to estimate the potential carbon dioxide reduction for a given allocation strategy. The conducted survey did not identify any difference between the ranking of reduction, avoidance and replacement strategies for our urban and medium-small urban areas. The contribution of research is reflected in the detailed assessment of generic scenarios and their application to real-world case studies. Also, the application of the presented methodology includes a ranking of transport strategies for mitigation of carbon dioxide emissions, assessment of strategies, setting budget priorities and developing estimates of mitigation potential.

Several research projects and models have been implemented in Europe to define and evaluate external costs in the transport sector starting from 2004 [13]. Among the most important are: HEATCO (Developing Harmonised European Approaches for Transport Costing and Project Assessment, 6th Framework Programme), CAFE CBA (Clean Air for Europe Programme, Cost Benefit Analysis of Air Quality), TREMOVE policy assessment model, ASSET (ASsessing SENSitiveness to Transport) GRACE (Generalisation of Research on Accounts and Cost Estimation, 6th Framework Programme). The most important results of these projects were summarised in the IMPACT project in 2008, which resulted in the Handbook on estimation of external costs in the transport sector [14]. The latest update related to external costs is provided in the Update of the Handbook on External Costs of Transport [15]. On the basis of manuals it is possible to calculate unit costs of pollutants in the form of €/tonne which take into account the negative consequences of traffic functioning such as: harmful impact on human health (mortality, morbidity), impact of emissions of harmful substances on facilities and materials, negative impact on the biosphere, detrimental effect on biodiversity and ecosystems, impact on the generation of greenhouse gases.

III. METHODOLOGY - QUANTIFICATION OF EXTERNAL EFFECTS

In order to more easily eliminate the negative consequences that accompany the entire transport process, it is necessary to note that external costs do not apply to the individual or direct user of the transport service when they occur, such as: Operating

costs of using the vehicle, Costs of own travel time and Costs of fees and charges in transport. External costs clearly cover all those costs that relate to society as a whole, and the following costs are then included: Accidents, Congestion, Air pollution, Climate change, Noise, Well-to-tank emissions and Habitat damage [16].

The internalization of external costs implies that such effects are part of the decision-making process of transport users. This can be done directly (by regulating both command and control measures) and indirectly (by providing better incentives for the transport of users, i.e. market instruments).

1. Air pollution

Air pollution caused by traffic activities leads to different types of external costs. The most important external costs are health costs due to cardiovascular and respiratory diseases caused by air pollutants. The most important air pollutants related to traffic are particles (PM₁₀, PM_{2.5}), Nitrogen Oxides (NO_x), Sulfur Dioxide (SO₂), Volatile organic compounds (VOC), and Ozone (O₃) as an indirect pollutant [17].

For road transport, the most important impact on costs is the emission standards for vehicles, which partly depend on the age of the vehicle. The emissions of road vehicles also depend on vehicle speed, fuel type and fuel combustion technology and exhaust gas treatment technologies, load factors, vehicle size, type of drive and geographical location of the road. The results of using a simulation model to estimate the fuel consumption of a light commercial vehicle in road traffic cycles are given in [18].

The quantification of external costs of pollutant emissions in road transport when transporting goods can be carried out depending on a particular category of vehicle and its emission class, i.e. euro standards. The costs of air pollution depend on the area through which a particular vehicle carries out the transport of goods. Costs are expressed in €c per vehicle-kilometer. The highest costs of air pollution are recorded passing through a city zone, while the lowest costs of air pollution are recorded on highways [15].

The key cost impacts for rail transport are: vehicle speed, load factors, a combination of power generation plants and the geographical location of plant installations. Calculating air pollution costs involves the use of linear functions and calculations that are included in the top-down model, which are also linear functions. Marginal costs of air pollution are approximately equal to the average cost of air pollution. External costs of air pollution from road and rail transport can be calculated on the basis of the form [19]:

$$C_m = \sum_{c,i} (V_{k_{m,c,i}} \cdot MC_{m,c,i}) \quad (1)$$

Where C - Air pollution costs per trip (€/trip), V_k - Vehicle kilometers (vkm/trip), MC - External marginal air pollution costs, i -Type of infrastructure (urban road, interurban road, motorway), c - Country and m -mode.

2. Accidents

Accident externalities represent the most important external costs of road transport [20]. These social costs include the costs of material damage, administrative costs, treatment costs, production losses and non-material costs (shortening life expectancy, suffering, pain, sadness, etc.) [21].

The most important impacts on costs in road transport, in addition to mileage, vehicle speed, road type, the characteristics of the driver (such as driver behaviour, experience, speed), the volume and speed of traffic, time of day (day/night) and interaction with weather conditions, the level of infrastructure maintenance, the degree of utilisation of the capacity of the infrastructure, and the level of segregation of road traffic lanes. The following form can be used to calculate the cost of traffic accidents in road transport [19]:

$$C_m = \sum_{c,i} (V_{k_{m,c,i}} \cdot MC_{m,c,i}) \quad (2)$$

Where C - Accident costs per trip (€/trip), V_k - Vehicle kilometers (vkm/trip), MC - External marginal air pollution costs, i -Type of infrastructure (urban road, interurban road, motorway), c - Country and m -mode.

According to German statistics, the share of fatal HGV (Heavy goods vehicle) accidents on highways is 50% of all fatal HGV accidents, which is very high compared to other countries. Combined with traffic flow data, this gives a marginal value in the case of an accident for highways that is higher than for other types of roads. In the original data in Switzerland, the share of motorway accidents was only 20% [15].

The main impacts on rail transport costs are traffic volumes, weather conditions, maintenance levels and level of segregation between systems, especially between road and rail transport and between different types of trains. To calculate the cost of accidents in railway transport one can use the form [19]:

$$C_m = \sum_c (V_{k_{m,c}} \cdot MC_{m,c}) \quad (3)$$

Where C - Accident costs per trip (€/trip), V_k - Vehicle kilometers (vkm/trip), MC - External marginal air pollution costs, c - Country and m -mode.

In the rail transport sectors of the EU, accidents are much less common than in road transport. Therefore, the cost estimation of accidents must be based on the average number of accidents in the past few years.

All incident costs can be considered as external, because the marginal costs are equal to average costs.

IV. ECONOMIC MEASURES AND INSTRUMENTS FOR SOLVING EXTERNAL EFFECT PROBLEMS

In order to preserve human health and natural resources of the environment, it is necessary to introduce certain measures aimed at reducing the negative impact of transport on the environment and human health. Economic instruments are a system of incentives, which are established with the aim of influencing the behaviour of economic entities as well as the decisions they make in order to protect and preserve environmental resources. In this way, environmental resources receive appropriate prices that will ensure their proper allocation and efficient and sustainable use. The application of economic instruments ensures that economic entities behave in a different way than before. Namely, they no longer treat the ecological good as free of charge and do not transfer the costs of endangering the environment to society, but take responsibility for them themselves. The European Environment Agency groups external instruments into five basic categories [22]: 1. Transferable permits, 2. Environmental taxes, 3. Environmental fees, 4. Environmental subsidies and incentives, and 5. Liability and compensation programmes.

1. Transferable permits are authorisations granting the right to emit pollutants. These permits are issued by the competent regulatory authority. The right to emit pollutants actually represents the amount of pollutants that an economic entity can produce. As a rule, these permits are sold to those who offer the highest price. The definition and introduction of the system of transferable licences is carried out in three steps: determining the total amount of quotas, allocation or distribution of licences to individual economic agents and trade in licences.
2. Environmental taxes are defined as mandatory payments imposed on products and processes that are harmful to the environment. These taxes have three basic functions: covering costs, encouraging behaviour change, and generating revenue. In the countries of the European Union, the following classification of environmental taxes is found:
 - tax on energy sources (on petrol, on diesel, on mineral oils, on heating oil, on kerosene, on natural gas, on electricity consumption and on CO₂),
 - tax on transport (on the registration and use of motor vehicles, on the import and sale of motor vehicles, on the use of roads and highways and on passengers in air traffic),
 - tax on pollution and resources (air pollution (SO₂, NO_x), water pollution, waste

(landfills), batteries, tires, available containers, plastic bags, pesticides and fertilisers, wastewater and industry), ozone pollution, nuclear energy and noise).

3. Environmental compensations are a mechanism created with the aim of partially or completely covering the costs of environmental services and mitigation measures such as wastewater treatment or waste disposal. Depending on the legislation, the taxpayer may pay these fees, may be exempted from them, or may be reimbursed, while a reduction in the payment of the fee may also be granted, if these funds are used to reduce environmental pollution.
4. Environmental subsidies and incentives include a set of government measures designed to stimulate new technologies, develop new markets for environmental products and services, encourage consumer behaviour change through eco-purchasing programmes and provide temporary support to businesses to ensure a higher level of environmental protection. Thus, in contrast to the taxes that are intended to punish polluters, subsidies are intended to influence the change in the behaviour of various actors in order to reduce pollution and improve the quality of the environment. Similar to environmental taxes, subsidies can be classified into four groups: energy-related subsidies, resource-related subsidies, pollution-related subsidies, and transport-related subsidies.
5. Liability and compensation programmes aim to provide adequate compensation for damage caused as a result of activities that are harmful to the environment and thus to provide funds for prevention and restoration.

Economic instruments to solve these problems have both positive and negative sides. The advantages of these instruments include: creation of additional state revenue, ensuring cost efficiency, ensuring dynamic efficiency through the operation of innovative activities in the long run, greater flexibility as they provide easier acceptance and adaptation of stakeholders to these instruments, and control of a large number of small and widespread pollution sources, which is especially characteristic of the transport sector, given that there are a large number of vehicles in it. The shortcomings of these instruments that are most often mentioned are: uncertainty when it comes to the appropriate level of duty, uncertainty about the time gap, uncertain and unstable revenues and effects on competitiveness.

V. RESULTS AND DISCUSSION OF RESEARCH - CALCULATION OF EXTERNAL COSTS OF RAILWAY AND ROAD FREIGHT TRANSPORT THROUGH THE REPUBLIC OF SERBIA

Negative external effects caused by daily activities of all modes of transport can only be expressed in the form of monetary units in order to assess their impact on both the individual and society as a whole.

The paper presents the calculation of external costs (air pollution and accidents) caused by different modes of transport (road and rail) between the border crossings of the Republic of Serbia with Hungary and North Macedonia, and on the route between the Kelebija and Preševo border crossings (**Fig. 1**).

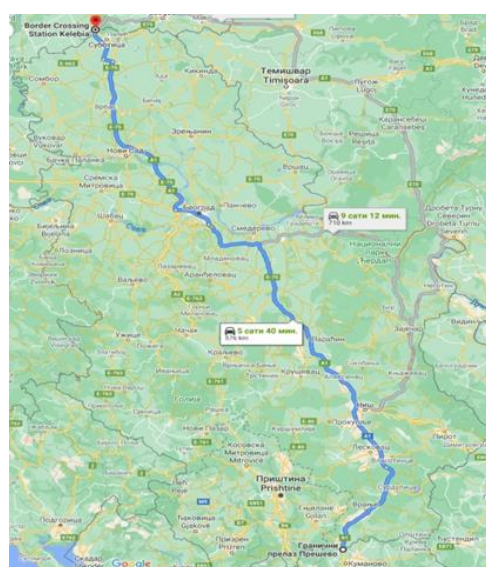


Figure 1. Movement path of the truck

The assumed cargo to be transported is 20 TEU (Twenty-foot Equivalent Unit) containers (adopted average weight is 10.5 t/TEU). The transport is considered for the following vehicle categories:

- Truck (Rigid HGV), Gross vehicle weight 24-40 t, Emission class Euro V,
- Train (Short train, Train weight 500 t, electrified and diesel).

The lengths of the routes of transport vehicles of different modes of transport through the Republic of Serbia are given in **Table 1**.

External costs caused by air pollution and traffic accidents for road and rail transport are calculated on the basis of equations (1) to (3) and adopted characteristics of transport vehicles [15, 19].

Table 1. Length of road and rail freight transport routes through the Republic of Serbia

	ROAD	RAIL
Republic of Serbia	BC Kelebija – BC Preševo = 576	BC Kelebija – Subotica – Novi Sad – Belgrade = 183 Belgrade – Niš = 241 Niš – Preševo – BC Preševo = 157
TOTAL (km)	576	581

Table 2. External costs caused by air pollution for road and rail transport

Freight transport through the Republic of Serbia				€
ROAD		Regional road	Highway	176.302
	km	34	543	
	Coefficient [23]	1.455		
RAIL		Electrified (diesel)		21.596
	km	581		
	Coefficient [23]	0.177		

Table 3. External costs caused by traffic accidents for road and rail transport

Freight transport through the Republic of Serbia				€
ROAD		Regional road	Highway	229.132
	km	34	543	
	Coefficient [23]	1.891		
RAIL		Electrified (diesel)		14.031
	km	581		
	Coefficient [23]	0.115		

Table 4. Total external costs caused by air pollution and traffic accidents for road and rail transport

	Freight transport through the Republic of Serbia		Total €
	Air pollution	Traffic accidents	
ROAD	176.302	229.132	405.434
RAIL	21.596	14.031	35.627

The obtained results are given in **Table 2** and **Table 3**, respectively.

If 20 TEU containers were transported through the Republic of Serbia by road or rail, external costs would amount to €405,434 and €35,627, respectively (**Table 4**). Regarding the comparison of the obtained results, the external costs of freight road transport are about 11 times higher than freight rail transport. When observing external costs caused by air pollution, freight road transport affects air quality about 8 times more than freight rail transport, while when it comes to external costs caused by traffic accidents, freight road transport affects traffic safety about 16 times more than rail freight transport.

From the obtained results, it can be concluded that the transport of 20 containers through the Republic of Serbia, from GP Kelebija to GP Preševo, from the point of view of external costs is

best done by rail, which has the least impact on air pollution and traffic accidents.

In Serbia, it is necessary to direct as much of the freight road transport to the rail transport in order to reduce the negative external effects. The next step would be the introduction of high-speed rail systems in the freight transport that are still in the making. The business idea of high-speed rail systems would be "Faster than trucks, cheaper than airplanes". The fact that Serbia aspires to become a member of the European Union and as such must follow the environmental protection measures implemented in the European Union (it would be desirable for it to at least start implementing them), only gives all this greater significance.

Freight transport is essential for economic activities and quality of life. Well-organized freight transport also contributes to sustainability and increased energy efficiency. In Serbia, the most appropriate strategies in dealing with the environment and problems of freight transport are:

- a) Mandatory legislation with appropriate enforcement measures, and
- b) Gradual dissemination and adoption of the best EU environmental practices.

Serbia is the only country among the pre-accession countries that applies the EU guidelines in the internalisation of external costs of road, rail, air, inland waterway, intermodal and multimodal transport.

VI. CONCLUSION

A particularly important strategic role of transport is reflected in its contribution to the opening of underdeveloped regions or economies and their integration into national, European and global economic flows. Transport is one of the most important factors in achieving the overall economic prosperity of a country and the economic well-being of its citizens. The increase in transport in the world economy has a number of shortcomings, such as congestion, increased noise and stress among traffic participants, environmental pollution by emissions, etc. Consideration and quantification of these effects were the first steps towards modelling alternative development scenarios and their costs, and later moving on to their internalisation. Both in the professional circles and in the public, it is understood that the costs of transport do not include only what the state or users pay for the transport service, but that they cover a much wider set of costs. They can be considered from the aspect of individual direct users, from the aspect of transport companies and from the aspect of society in two ways, as a direct user of transport services in some cases, when at some level public transport is required and paid, and as an indirect user of transport, those on which the traffic system operates in any way. The principle of internalisation of external costs involves adding external costs to the individual costs of polluters. The internalisation of external costs has the effect of increasing the price of pollutant products and reducing the demand for their products. It is not only important to determine the value or type of external transport costs, but it is also necessary to dedicate oneself to solving the problems that arise from that. Involving as many stakeholders as possible, as well as using a wide range of instruments, is crucial to solving these problems.

REFERENCES

- [1] N. Petrović, N. Bojović, M. Petrović, V. Jovanović, A study of the environmental Kuznets curve for transport greenhouse gas emissions in the European Union, *Facta Universitatis Series: Mechanical Engineering* 18 (3) (2020) pp. 513 – 524. <https://doi.org/10.22190/FUME171212010P>
- [2] A. Ljungberg, Marginal cost-pricing in the Swedish transport sector – An efficient and sustainable way of funding local and regional public transport in the future?, *Research in Transport Economics* 59 (2016) pp. 159-166. <https://doi.org/10.1016/j.retrec.2016.05.005>
- [3] Handbook on the external costs of transport – Version 2019, European Commission, Brussels (2019) [cited 2022-02-15] <https://op.europa.eu/en/publication-detail/-/publication/9781f65f-8448-11ea-bf12-01aa75ed71a1>

A comparative analysis of the representative negative external costs, which arise from freight road and rail transport, should lead to redirecting users or decision makers to an acceptable mode of transport. An example of transporting 20 TEU containers through Republic of Serbia and quantifying the use of two transport modes shows that the highest external costs are incurred when using trucks, while the lowest external costs are caused by rail transport.

When choosing a particular mode of transport for the transport of goods, it is necessary to consider all costs, as far as possible, and not only the cost of transport services.

ACKNOWLEDGEMENT

This research was financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract No. 451-03-68/2022-14/ 200109).

AUTHOR CONTRIBUTIONS

- N. Petrović:** Conceptualization and Supervising.
- V. Jovanović:** Conceptualization and Experiments.
- B. Nikolić:** Writing and Reviews.
- J. Pavlović:** Writing and Reviews.
- J. Mihajlović:** 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>
- B. Nikolić** <https://orcid.org/0000-0002-9694-6719>
- J. Pavlović** <https://orcid.org/0000-0002-8825-7307>
- J. Mihajlović** <https://orcid.org/0000-0002-9787-0226>

- [4] Republic Bureau of Statistics, Statistical Calendar of the Republic of Serbia-2021, online edition, Belgrade (2021) [cited 2022-02-16] <https://publikacije.stat.gov.rs/G2021/pdf/G202117014.pdf>
- [5] Z. Bokor, Cost Calculation in Transport Companies, *Acta Technica Jaurinensis* 5 (3) (2012) pp. 253-262. <https://acta.sze.hu/index.php/acta/article/view/61>
- [6] L. Lazarević, M. Kovačević, Z. Popović, Rail traffic volume estimation based on world development indicators, *Facta Universitatis Series: Mechanical Engineering* 13 (2) (2015) pp. 133–141.
- [7] F. Petroa, V. Konečnyá, Calculation of emissions from transport services and their use for the internalisation of external costs in road transport, *Procedia Engineering* 192 (2017) pp. 677 – 682. <https://doi.org/10.1016/j.proeng.2017.06.117>
- [8] E. Christen, B. Meinhart, F. Sinabell, G. Streicher, External Costs of Freight Transport - Relevance and Implications of Internalization at the European Level, *SUERF Policy Brief* 221 (2021) https://www.suerf.org/docx/f_35c0435bac5b49fc667bd23a5c49feal_35885_suerf.pdf
- [9] J. Cravioto, E. Yamasue, H. Okumura, N. K. Ishihara, Road transport externalities in Mexico: Estimates and international comparisons, *Transport Policy* 30 (2013) pp. 63-76. <https://doi.org/10.1016/j.tranpol.2013.08.004>
- [10] A. Agarwal, B. Kickhoefer, The correlation of externalities in marginal cost pricing: lessons learned from a real-world case study, *Transportation* 45 (2018) pp. 849–873. <https://doi.org/10.1007/s11116-016-9753-z>
- [11] C. Ortolani, A. Persona, F. Sgarbossa, Modeling external transport costs in distribution networks, *POMS 20th Annual Conference Orlando, Florida U.S.A.* (2009) May 1 - 4.
- [12] J. R. Javid, A. Nejat, K. Hayhoe, Selection of CO₂ mitigation strategies for road transportation in the United States using a multi-criteria approach, *Renewable and Sustainable Energy Reviews* 38 (2014) pp. 960-972. <https://doi.org/10.1016/j.rser.2014.07.005>
- [13] I. Ivković, Research of performances of the Compressed Natural Gas powered bus in terms of safety and environmental influence, Ph.D. thesis, University of Belgrade, Faculty of Transport and Traffic Engineering (2012). <https://nardus.mpn.gov.rs/handle/123456789/2665>
- [14] M. Maibach, C. Schreyer, D. Sutter, H. P. Essen, B. H. Boon, R. Smokers, A. Schroten, C. Doll, B. Pawlowska, M. Bak, Handbook on estimation of external costs in the transport sector - Produced within the study Internalisation Measures and Policies for All external Cost of Transport (IMPACT), CE Delft (2008).
- [15] A. Korzhenevych, N. Dehnen, J. Bröcker, M. Holtkamp, H. Meier, G. Gibson, A. Varma, V. Cox, Update of the Handbook on External Costs of Transport, European Commission – DG Mobility and Transport, Ricardo –AEA, Issue 1 (2014).
- [16] H. Essen, E. Andrew, D. Sutter, et al., Directorate-General for Mobility and Transport, Sustainable transport infrastructure charging and internalisation of transport externalities: main findings, European Commission, Publications Office (2019). [cited 2022-02-17] <https://data.europa.eu/doi/10.2832/004905>
- [17] N. Petrović, B. Krstić, J. Petrović, Evaluation of freight transport modes based on external costs, Proceedings of 14th International conference on accomplishments in mechanical and industrial engineering, University of Banja Luka, Faculty of Mechanical Engineering (2019) pp. 579-584.
- [18] A. Kolin, S. E. Silantsev, P. Rogov, M. E. Gnenik, Methods and simulation to reduce fuel consumption in driving cycles for category N1 motor vehicles. *Acta Technica Jaurinensis* 14(4) (2021) pp. 477–487. <https://doi.org/10.14513/actatechjaur.00593>
- [19] A. Schroten, H. Essen, R. Anthes, External Cost Calculator, Methodology report, CE Delft, Delft (2011).
- [20] M. Dementyeva, P. R. Koster, E. T. Verhoef, Regulation of road accident externalities when insurance companies have market power, *Journal of Urban Economics* 86 (2015) pp. 1-8. <https://doi.org/10.1016/j.jue.2014.11.001>
- [21] H. Essen, A. Schroten, M. Otten, D. Sutter, C. Schreyer, C. Zandonella, M. Maibach, C. Doll, External Costs of Transport in Europe - Update Study for 2008, CE Delft, Delft (2011).
- [22] S. M. Kaplanović, Internalisation of external costs for the purpose of providing sustainable development of road transport, Ph.D. thesis (in Serbian), University of Belgrade, Faculty of economics (2012). <https://nardus.mpn.gov.rs/bitstream/handle/123456789/2214/Disertacija.pdf?sequence=4&isAllowed=y>
- [23] Adriatic - Danube - Black Sea multimodal platform, WP 6: “ADB and Green Transport”, Project title: Autonomous Region of Friuli Venezia Giulia (2013) [cited 2022-02-17] <http://amministrazioneaperta.regione.fvg.it/opendata/p/files/50f9404c-e25d-4e9d-8ac5-5e52b25ecb6a>



Pave-Ut—Hungarian environmental load application

Seoyoung Cho^{1,*}, Csaba Tóth¹

¹Department of Highway and Railway Engineering, Budapest University of Technology and Economics
Műegyetem rkp. 3, 1111, Budapest, Hungary

*e-mail: cho.seoyoung@emk.bme.hu

Submitted: 21/04/2022 Accepted: 04/06/2022 Published online: 08/06/2022

Abstract: Structural analysis of road pavement structures is necessary to reduce material waste in cross-sectional design and to accompany optimal feeling of use within the service life. As a result of many studies, it is known that the effect of environmental load on roads is significant. Since environmental loads have regional characteristics, this study focused on the application of environmental loads in structural analysis in the light of regional characteristics in Hungary. In this study, a structural analysis program called Pave-Ut, which is based on the finite element analysis method, was developed. Among the major environmental factors, stiffnesses of the temperature influence on asphalt layer and water content change influence on subgrade are featured. Pave-Ut was verified through BISAR and ALVA, which are codes for research. An error of about 10% occurred, which will be reinforced in future studies. Finally, a comparative analysis was carried out in a case in which the asphalt temperature profile for each layer was reflected and a case in which the stiffness of the subgrade was reflected in the monthly average rainfall, and these were compared to a case in which these characteristics were not taken into account.

Keywords: asphalt material property programming; dynamic modulus at specific depth; dynamic modulus; Hungarian asphalt temperature prediction; pavement structural analysis

I. INTRODUCTION

Structural analysis in road pavement design is a dynamic prediction process for the horizontal or vertical strain in the surface layer, base layer, and subgrade [1]. The predicted horizontal or vertical strain becomes the main variable in serviceability analysis.

The aim of this study is to investigate the influence of the climatic indexes on pavement system, especially the temperature and water content. The numerical model for the temperature distribution inside of the asphalt layer is prepared and the application of water content change on the subgrade modulus variation is carried out. The result of this study shows that the temperature effect on asphalt layer's stiffness and structural responses with consideration of this change is distinguishably different from the simple design condition with fixed stiffness value which does not consider such change. Whereas the subgrade modulus variation does not show big difference in response.

1. Literature Review

A design input variable provides the basic data for estimating the optimal thickness of pavement. Design input variables are used to predict the mechanical behavior and utility of a preliminary pavement section during the service period [2, 3]. For design input variables, the environmental characteristics include the internal temperature of the asphalt concrete pavement, the water content of the sub-base and subgrade, and the freezing index. The internal temperature of an asphalt concrete pavement layer is determined in the form of the distribution with respect to depth by using the temperature prediction model of the road pavement structure design based on the atmospheric temperature [4, 5]. The asphalt concrete temperature is known to be influenced by some weather indexes. Chao et al. investigated the effects of those indexes on the pavement temperature. The study revealed that the air temperature and relative humidity had a great influence on the asphaltic temperature, whereas cloud cover, wind speed, and precipitation had weaker influences [2]. The sub-base and subgrade moisture contents are determined by using the water content prediction model of the road pavement

structure design with the average monthly atmospheric temperature, the average monthly accumulated precipitation, and the grain size characteristics of the sub-structure material.

Asphalt is a petroleum compound that left over after crude oil refining which is sticky, black viscous liquid or semi-solid. Asphalt exhibits a very complex rheological behavior, which mainly depends on temperature, traffic load, and application rate. At low temperatures and low load application rates, asphalt acts as an elastic solid, while at high temperatures and longer load rates, it acts as a viscous liquid. At medium temperatures, it is characterized as a viscoelastic fluid [3]. To determine the importance of the effect of temperature on asphalt's durability, many studies were carried out to estimate the temperature [8, 9].

In the design stage of a road, the viscosity as a function of the temperature is important due to the pavement's aging and distresses caused by climate change. In NCHRP Project 1-40D, a part of the Mechanistic-Empirical Pavement Design Guide, Bari and Witzczak proposed a set of predictive models for the viscosity and modulus of asphalt binders [4].

Environmental and climatic factors that affect the behavioral characteristics of asphalt concrete pavement structures are largely divided into changes in the physical properties of the asphalt concrete layer due to changes in atmospheric and internal temperature and changes in the properties of the subgrade, granular layer, and sub-base layer due to seasonal influences. It is important in terms of road engineering to predict these characteristics in advance and to reflect them in the structural capabilities.

Temperature sensitivity plays an important role in understanding asphalt pavement failure and indicates how quickly asphalt properties change over time in terms of indicators such as penetration index [5]. There are many ways to detect the road surface. In this study, direct temperature measurement was performed using a thermometer. Another study presented road background segmentation based on watersheds [6].

Efforts have been made to accurately predict the temperature in the structure of asphalt concrete pavement. The calculation of the pavement temperature with the thermal properties of pavements started with the work of Barber. A maximum temperature estimation was predicted with observed weather data. This work made a correlation of locally limited and time-limited observed data with a standard weather report [7]. External factors, such as insolation, atmospheric temperature, wind speed, precipitation, cloud coverage, and subgrade water content, all influence the temperature of the pavement. The internal

conditions of the binders, as well as the types of binders, have an impact. Dickinson incorporated external meteorological conditions into the temperature estimation of asphaltic pavements [8].

Saas created a road condition model that was integrated into the Danish Meteorological Institute's autonomous road temperature prediction system. The findings of this investigation revealed that the temperature forecast sensitivity is highly dependent on atmospheric meteorological data [9]. At the request of the Canadian weather center, a system named METRO was built in Canada for scientific purposes. This system has the advantage of being able to use weather and road temperature observation values from the road meteorological information system, as well as weather data projected by the Canadian weather center's own model. It can also explain the state of the collection of water on the road surface and whether it is liquid or solid [10].

The temperature equivalency factor, which is a function of the type of distress, the structure, and the material characteristics, was suggested. This was applied to convert in-service traffic loading into its equivalent at the reference time [11].

To prepare the temperature profile, the heat transfer theory and heat balance at the surface were considered [12]. The internal temperature distribution could be predicted from the surface temperature of asphalt concrete [8]. The prediction of the surface temperature of asphalt concrete is the result of the heat exchange balance between the asphalt concrete and the atmospheric temperature based on the thermal equilibrium equation. From the net energy balance of the pavement structure, Solaimanian and Kennedy proposed a quadratic equation for the prediction of the maximum pavement surface temperature [13].

Based on the BELLS model, a temperature prediction equation at the midpoint of the asphalt was proposed in Tennessee. With this predicted temperature, the expression for inversely calculating the modulus of asphalt concrete was expressed as an exponential function [14].

Witzczak set the temperature of the one-third depth point of an asphalt concrete layer that was calculated using the atmospheric temperature as the representative temperature and proposed an equation for calculating the elastic modulus of the asphalt concrete layer using this. In addition, Witzczak proposed an equation for obtaining the modulus of elasticity at an arbitrary temperature based on the modulus of elasticity of the asphalt concrete layer at 25 °C [15]. Currently, the American Asphalt Association sets the temperature at one-third of the asphalt concrete layer's depth as the representative temperature and sets the standard temperature at 21 °C. The modulus of elasticity is corrected and used.

Dickinson predicted the internal temperatures of pavement structures by developing a finite difference temperature prediction program that used Australia's solar radiation, average monthly maximum temperature, and meteorological data, such as sunrise and sunset times, as input data [8]. As part of the NCHRP 1-37A project, an empirical–mechanical design method for asphalt properties was established to predict the viscoelasticity and complex shear modulus of asphalt binders based on an extensive binder characterization database [4].

When the wheel load is applied, the element within the lower part of the pavement system is confined not only by the pavement system above it, but also by the wheel load's repeated deviator stress. As the number of repetitions of loads rises, one piece of the package's bottom will repeat elastic deformation and recovery, accumulating plastic deformation. The recovery strain rate becomes more prominent than the plastic deformation as the number of repetitive loads rises, and the slope of the stress–strain curve is characterized as the recovery elasticity factor (Fig. 1). Under repeated stress, the subgrade exhibits nonlinearity, which is a deformation that is recoverable from some fraction of the plastic deformation [16].

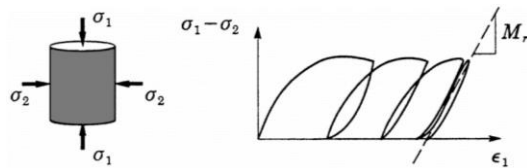


Figure 1. Triaxial test and resilient behavior of granular materials.

The stiffness of a material with respect to a resilient response can be found with equation (1):

$$M_r = \frac{\sigma_d}{\epsilon_r} \quad (1)$$

This implies that a resilient modulus (M_r) is expressed as the deviatoric stress (σ_d) divided by the recoverable strain (ϵ_r).

In an effort to describe the resilient behavior of the subgrade, the earliest model of $K - \theta$ (equation 2) arose from Hicks and Monismith [17].

$$M_r = k\theta^n \quad (2)$$

The mean pressure acting on the specimen (θ , bulk stress) is raised to a power, n , in this simple model. Many modifications have been devised by specialists from this simple design. Uzan discovered that deviator stress had a non-negligible effect on the resilient modulus and included a deviator term in his model [18]. Later, Witczak built the Witczak–Uzan model by substituting the deviator component from the Uzan model into octahedral stress [19]. This model is extensively utilized in three-dimensional

FE modeling, since the octahedral stress in the model is a three-axis term (equation 3).

$$M_r = k_1 P_a \left(\frac{\theta}{P_a} \right)^{k_2} \left(\frac{\tau_{oct}}{P_a} \right)^{k_3} \quad (3)$$

where k_1 , k_2 , and k_3 are regression coefficients from triaxial test, τ_{oct} is the octahedral shear stress ($= \frac{1}{3} \sqrt{(\sigma_1 - \sigma_2)^2 + (\sigma_1 - \sigma_3)^2 + (\sigma_2 - \sigma_3)^2}$), and P_a is the atmospheric pressure.

The material properties of the pavement substructure are the modulus of elasticity and Poisson's ratio. For the empirical model for determining the design input variables, the volumetric stress model is applied.

$$E = k_1 + k_2 \cdot \theta \quad (4)$$

where E is the elastic modulus, θ is the bulk stress ($= \sigma_1 + \sigma_2 + \sigma_3$) [kPa], and k_1 and k_2 are model coefficients (Table 1).

Table 1. Range of the properties of a crushed stone base material [38].

Index	Range
Elastic modulus	$100 \leq E \leq 600$
k_1	$80 \leq k_1 \leq 270$
k_2	$0.1 \leq k_2 \leq 0.6$

Whether the road structure satisfies the safety conditions of the specification for various load conditions should be reviewed. In general, satisfaction is checked through experiments or numerical analyses. The method of design verification through numerical analysis is generally applied to the design formula presented in the design regulations or to the method of numerical analysis using finite element analysis. Depending on the physical characteristics of the applied load, various stability evaluation methods, such as structural analysis, flow analysis, vibration analysis, and fatigue life evaluation, are applied.

For mechanical analysis, tools for structural analysis are essential. In the case of road pavement using only elastic materials, structural analysis can even be performed well with the multi-layer elasticity theory, but as the complexity of materials and interlayer boundary conditions increases, the use of numerical methods, such as finite element analysis, is essential.

Duncan and his colleagues prepared the axisymmetric structural modeling of asphalt concrete and suggested the boundary condition location at a depth of 18 radii for the bottom and vertical boundaries at 12 radii from the center. The Gonzales bypass process was used to reflect the granular layer's resilient modulus. An emphasis was made on the fact that the variation of the resilient

modulus in the vertical direction had to be adapted in the model. In addition, the different temperatures were set at each node of the asphalt concrete to show the nonlinearity [16]. It appeared that the applied load in the prediction of deflection could be treated as static loading according to the results of laboratory tests. This is thanks to the small influence of an inertial effect [20].

Computer programs that are used to solve the boundary value problem of a multilayered pavement system fall into two groups: finite element theory and elastic layer theory [21]. ALVA calculates the structural responses of pavement based on Burmister's layered elastic theory [22].

Nowadays, finite-element-based programs are used widely thanks to their versatility, whereas layered elastic analysis based on closed-form solutions is not sufficient to characterize the materials used [23]. Currently, various codes for finite element analysis have been developed, and there are commercial products such as ANSYS and ABAQUS that are manufactured for general use, as well as research codes, such as ILLIPAVE and EAPAVE, which are manufactured exclusively for road pavement. ILLIPAVE addresses the axisymmetric problem and applies the K- θ method for granular materials [24].

Commercial code has the advantage of having a variety of material models, but has the problem of being expensive to use. Research code also has a disadvantage in that it is almost impossible to add elements or materials because access to the source-code level is limited. Therefore, to solve these problems, we developed a finite element code. To apply the nonlinear properties of an asphalt concrete layer to the structural analysis, the method of subdividing the layer and applying the elastic modulus was adopted.

The finite element analysis code developed for road pavement was named Pave-Ut. Partially using both C# and MATLAB, the code is for an axisymmetric problem. In this study, an introduction of the axial symmetry analysis method implemented in Pave-Ut, the formulation of the calculation of the input material properties for the analysis, and the process of the numerical verification of the developed code are additionally described.

The biggest reason to predict the temperature of the road pavement and apply it to the structural analysis is to predict the bearing capacity more accurately. This is the basis of life cycle cost in construction [25]. Also, according to a numerical study conducted previously, it was found that using only one stiffness for composite structures was insufficient [26]. Therefore, in this study, various stiffnesses considering the temperature distribution were applied to the asphalt mixed layer modeling.

II. RESEARCH METHOD

In this study, research was conducted on the prediction of the temperature of existing asphalt pavement according to time and depth by using the temperature measurement data at the actual site. A temperature sensor for each depth was buried in the center of Budapest, and the surface and internal temperatures of the pavement were periodically measured every 15 minutes. Structural analysis based on the finite element analysis method was performed by analyzing these field data and applying the internal asphalt temperatures estimated using the temperature prediction formula used in Hungary to the prediction of the stiffness of the asphalt concrete layer. Based on the results of this study, a method for applying environmental loads suitable for the climate and environmental conditions of Hungary is suggested.

2. Material Properties

Based on the predictive Witzcak model, a program that can calculate the dynamic modulus at a specific depth was prepared. In addition, a predictive model for the temperature at a specified depth of asphalt was introduced for the Hungarian climate, which would be an input for the program. This calculator program was built with C#, an object-oriented computer language with the aim of free usability for anybody while also saving computational capacity.

A. Asphalt Concrete Layer

Internal Temperature

The internal temperature distribution can be predicted from the surface temperature of asphalt concrete [8]. The prediction of the surface temperature of asphalt concrete is the result of the heat exchange balance between the asphalt concrete and the atmospheric temperature based on the thermal equilibrium equation [13].

The solar energy from the sun on the surface of the pavement at any time is [27]:

$$I(t) = \frac{2S}{t_1} \sin^2 \frac{\pi t}{t_1} \quad (5)$$

S : total insolation for a day (Wh/m^2); t_1 : time (h , set as 0 at 1 hour before sunrise); t : time.

At any time on the surface of the pavement, the convective energy between the surface of the pavement and the atmosphere is:

$$E(t) = h_c [T_a(t) - T_s(t)] = h\Delta T(t) \quad (6)$$

h_c : coefficient of surface thermal transfer ($W/m^2 \cdot K$); T_a : air temperature (K); T_s : surface temperature (K).

Thus, the energy flux can be stated as:

$$\gamma I(t) - E(t) \quad (7)$$

γ : absorptivity of the surface for solar radiation.

A quadratic equation for the maximum and minimum pavement temperatures that considers the latitude was suggested by [13] based on the energy balance.

$$q_{net} = q_s + q_a - q_c - q_k - q_r = 0 \quad (8)$$

q_s : direct solar radiation; q_a : atmospheric radiation; q_c : convection energy; q_k : conduction energy; q_r : radiation energy emitted from the surface.

From the net energy balance of the pavement structure, Solaimanian and Kennedy proposed the quadratic equation for the prediction of the maximum pavement surface temperature [13].

$$R_0 \cdot \alpha_1 \tau_a^{\frac{1}{\cos z}} \cos z + \varepsilon_a \sigma T_a^4 + h_c(T_s - T_a) - \frac{k}{x}(T_s - T_x) - \varepsilon \sigma T_s^4 = 0 \quad (9)$$

R_0 : solar constant (1367 W/m²); α_1 : solar absorptivity (asphalt concrete: 0.85–0.93); τ_a : transmission coefficient (clear day = 0.81, cloudy day = 0.62); z : Zenith angle; ε_a : coefficient of atmospheric radiation; σ : Stefan–Boltzman constant (5.67 × 10⁻⁸ W/m²°K⁴); k : thermal conductivity (1.36 W/m°C); T_x : temperature at depth x .

The solution of this equation using the finite difference method is developed as the following equation. In this study, the finite difference method is used to solve the problem, and the space of the asphalt concrete layer is divided into the finite spatial element Δx and time Δt . i is the location of the time node, and m is the location of the spatial node.

The change in temperature with time on the surface (difference equation) is:

$$\gamma I(t) + h(T_a - T_o^i) + \kappa \frac{T_1^i - T_o^i}{\Delta x} = \rho C_v \frac{\Delta x T_o^{i+1} - T_o^i}{2 \Delta t} \quad (10)$$

ρ : density (2.24 t/m³); κ : thermal conductivity (W/m°K); h : specific heat (840 Ws/kg°K).

The estimation of the temperature at the bottom of the asphalt layer, which is calculated with energy conservation, is:

$$\kappa \frac{T_1^i - T_o^i}{\Delta x} = \rho C_v \frac{\Delta x T_o^{i+1} - T_o^i}{2 \Delta t} \quad (11)$$

With the boundary conditions described in equations 10 and 11, the temperature inside the pavement can be estimated.

The finite difference method is a method of solving a differential equation as an algebraic differential equation under general boundary

conditions, and the temperature of the center can be known [28]. The discretization of the space and time is performed with the mesh size of $\Delta x = 0.2$ and $\Delta t = 0.2$. In this study, the solution of the temperature distribution is calculated with the Crank–Nicolson implicit method (equation 12).

$$\begin{aligned} -\left(\frac{\alpha \Delta t}{\Delta x^2}\right) u_{i-1,j+1} + 2\left(\frac{\alpha \Delta t}{\Delta x^2}\right) u_{i,j+1} \\ -\left(\frac{\alpha \Delta t}{\Delta x^2}\right) u_{i+1,j+1} \\ = \left(\frac{\alpha \Delta t}{\Delta x^2}\right) u_{i-1,j} \\ + 2\left(\frac{\alpha \Delta t}{\Delta x^2}\right) u_{i,j} \\ + \left(\frac{\alpha \Delta t}{\Delta x^2}\right) u_{i+1,j} \end{aligned} \quad (12)$$

Based on the finite difference equation described above (equation 12), a Pave-Ut module was built with Matlab. The internal temperature prediction results will be applied to the further structural analysis of the pavement system as an input.

Witzcak modulus prediction model

A dynamic modulus prediction model that was prepared by Witzcak for the National Cooperative Highway Research Program (NCHRP) 1-37A is implemented in Pave-Ut [29].

$$\begin{aligned} \log_{10}|E^*| = & -1.249937 + 0.02923p_{200} \\ & - 0.001767(p_{200})^2 \\ & - 0.002841p_4 \\ & - 0.05809V_a - 0.082208 \frac{V_{eff}}{V_{eff} + V_a} \\ & + \frac{3.971877 - 0.0021p_4 + 0.003958p_{3/8} - 0.000017(p_{3/8})^2 + 0.00547p_{3/4}}{1 + \exp(-0.603313 - 0.313351 \log f - 0.393532 \log \eta)} \end{aligned} \quad (13)$$

where p_{200} is the percentage of aggregate passing a #200 sieve, p_4 is the percentage of aggregate retained in a #4 sieve, $p_{3/8}$ is the percentage of aggregate retained in a 3/8 inch sieve, $p_{3/4}$ is the percentage of aggregate retained in a 3/4 inch sieve, V_a is the percentage of air voids, V_{beff} is the percentage of effective asphalt content, f is the loading frequency (Hz), and η is the binder viscosity at the temperature of interest.

In this study, as an input of the dynamic modulus prediction model (equation 13), the viscosity (η) at a temperature is used according to the value predicted by Bari and Witzcak [4].

$$\log \log \eta_{f_s,T} = A' + VTS'T_R \quad (14)$$

$$A' = 0.9699f_s^{-0.0527} \times A \quad (15)$$

$$VTS' = 0.9668f_s^{-0.0575} \times VTS \quad (16)$$

where $\eta_{f_s,T}$ is the viscosity of the asphalt binder as a function of both the loading frequency (f_s) and the temperature (T_R , Rankine scale; cP), A is a regression intercept from the conventional ASTM $A_i - VTS_i$, VTS is the slope from the conventional ASTM $A_i - VTS_i$, and A' and VTS' are adjusted values of A and VTS .

B. Subgrade

In all design stages, the elastic modulus of the subgrade should consider long-term and short-term changes in the water content [30].

The resilient modulus prediction model for the pavement considering the water content is as follows [31] [32].

$$\log\left(\frac{M_R}{M_{Ropt}}\right) = a + \frac{b - a}{1 + \exp\left(\ln\left(-\frac{b}{a}\right) + k_m \cdot (S - S_{opt})\right)} \quad (17)$$

where M_R/M_{Ropt} is the resilient modulus ratio, a is the minimum of $\log(M_R/M_{Ropt})$, b is the maximum of $\log(M_R/M_{Ropt})$, k_m is the regression parameter, and $(S - S_{opt})$ is the variation in the degree of saturation expressed in decimals.

This model was updated later and applied into Korean road design specifications [33].

The subgrade water content is determined by a model that includes the monthly average temperature, monthly accumulated precipitation, and soil distribution characteristics (Table 2). The model in this specification suggests using the following water content inputs.

$$E = k_1 \theta^{k_2} \sigma_d^{k_3} 10^{k_w(\omega - \omega_{opt})} \quad (18)$$

$k_1, k_2,$ and k_3 are model coefficients, ω_{opt} is the optimal water content, ω is the water content, and k_w is a model coefficient that reflects the water content, which is -0.1417 (for coarse-grain material) or -0.0574 (for fine-grain material). E_{opt} is the estimated subgrade modulus in an optimal water

content condition. $prec$ is the monthly precipitation (mm), and P_{200} is the #200 sieve passing ratio.

AASHTO guides the resilient modulus of the subgrade, which takes into account the deviation of the water content due to seasonal fluctuation [34]. Equation (19) is a generalized estimation of the resilient modulus based on AASHTO MEPDG, and the M_r determination test method is based on AASHTO T 307 or NCHRP 1-28A [30].

$$M_r = k_1 p_a \left(\frac{\theta}{P_a}\right)^{k_2} \left(\frac{\tau_{oct}}{P_a} + 1\right)^{k_3} \quad (19)$$

where P_a is the normalizing stress and $k_1, k_2,$ and k_3 are model coefficients.

3. Numerical Method

The key concept of a program using the finite element method is to build finite elements. There is a huge number of libraries of finite elements in commercial and scientific finite element programs. Thus, the study of the existing codes and their employment with modifications are important [35].

The purpose of this study is to prepare a calculator for the asphalt modulus at a specific depth (Fig. 2). To fulfill this purpose, we take the temperature as a function of the asphalt mix depth, which is the input of the binder viscosity (equations 14-16).

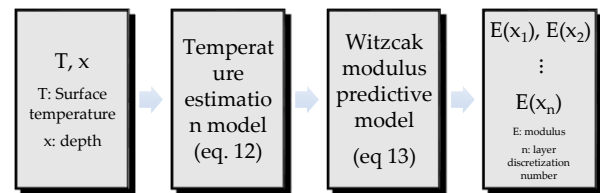
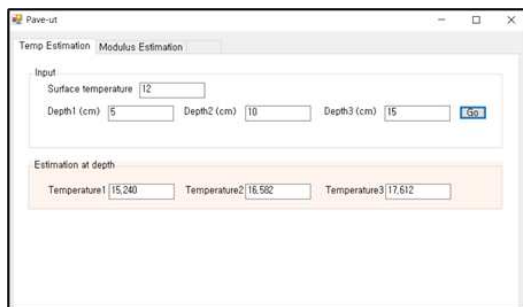
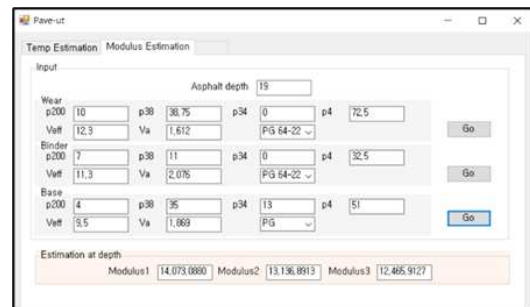


Figure 2. Scheme of modulus calculator.



(a)



(b)

Figure 3. Graphical user interface of Pav-Ut: (a) Temperature estimation part; (b) Modulus estimation part.

Table 2. Subgrade water content estimation model [3].

Region	Subgrade water content estimation model (ω)
Northern part	$\omega = 23.54759 + 0.15216 \times temp + 0.00070721 \times prec + 0.17990 \times P_{200}$
Southern part	$\omega = 21.84699 + 0.09598 \times temp + 0.00064287 \times prec + 0.9130 \times P_{200}$

The program first gets the surface temperature of the pavement and then estimates the temperature at the desired depth. Later, the user inputs the asphalt data; the program starts to match the input data and built-in material data, and then calculates the corresponding modulus.

Fig. 3 depicts the graphical user interface of the calculator. The program was made with two tabs; the first tab displays the temperature estimation and the second displays the dynamic modulus estimation.

The calculation time required for analysis increases in the order of frame element, plate element, and solid element [36]. A study indicated that the 3D analysis did not display major differences from the 2D axisymmetric analysis for predictions of the structural responses of pavement [37].

Displacements occur as a result of loading in the r and z directions, and they are denoted by u and z (Fig. 4).

Table 3. Pavement parameters used in the analysis

Index	Material properties	
	Stiffness [MPa]	Poisson's ratio
Asphalt	260	0.35
Base	500	0.4
Subgrade	Inf.	0.45

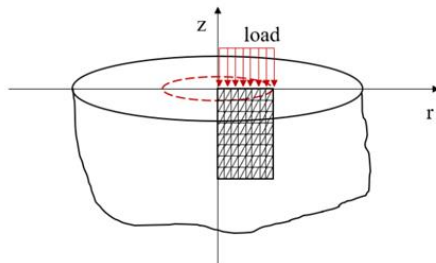


Figure 4. Axisymmetric finite element formulation.

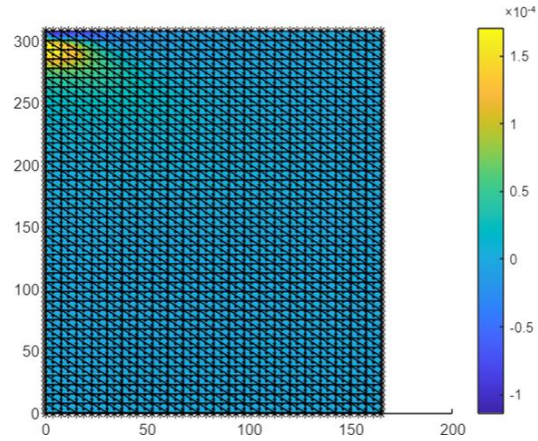
$$\begin{Bmatrix} \sigma_{rr} \\ \sigma_{zz} \\ \sigma_{\theta} \\ \tau_{rz} \end{Bmatrix} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & \nu & 0 \\ \nu & 1-\nu & \nu & 0 \\ \nu & \nu & 1-\nu & 0 \\ 0 & 0 & 0 & \frac{1-2\nu}{2} \end{bmatrix} \begin{Bmatrix} \epsilon_{rr} \\ \epsilon_{zz} \\ \epsilon_{\theta} \\ \gamma_{rz} \end{Bmatrix} \quad (20)$$

where σ_{rr} , σ_{zz} , σ_{θ} , and τ_{rz} are stresses in each direction, E is the stiffness, ν is Poisson's ratio, and ϵ_{rr} , ϵ_{zz} , ϵ_{θ} , and γ_{rz} are strains in each direction.

Equation 20 describes the structural responses in constitutive equation form that designers are interested in. The pavement structure has symmetricity in its geometry, applied loadings, and

the material properties. Thus, solving the axisymmetric problem with cylindrical coordinates is possible. Here, the tangential strain (ϵ_{θ}) depends only on the r direction.

Pave-Ut employs a six-node triangular element with full-bond interlayers and three Gaussian points for the calculations. The mesh size and boundary condition settings are standardized to avoid program crashes. Meshes are restrained in the horizontal direction, and the meshes in the bottom part are fixed



```

nnd = 0;
k = 0;
for i = 1:NXE
    for j = 1:NYE
        shapefunction;
        %
        geom(n1,:) = [(i-1)*dxx - X (j-1)*dyy - Y];
        geom(n2,:) = [(2*i-1)/2*dxx - X (j-1)*dyy - Y];
        geom(n3,:) = [i*dxx - X (j-1)*dyy - Y];
        geom(n4,:) = [(i-1)*dxx - X ((2*j-1)/2)*dyy - Y];
        geom(n5,:) = [(2*i-1)/2*dxx - X ((2*j-1)/2)*dyy - Y];
        geom(n6,:) = [i*dxx - X ((2*j-1)/2)*dyy - Y];
        geom(n7,:) = [(i-1)*dxx - X j*dyy - Y];
        geom(n8,:) = [(2*i-1)/2*dxx - X j*dyy - Y];
        geom(n9,:) = [i*dxx - X j*dyy - Y];
        %
        nel = 2*k;
        n = nel + 1;
        connection(nel,:) = [n1 n2 n3 n5 n7 n4];
        connection(nel,:) = [n3 n6 n9 n8 n7 n5];
        max_n = max([n1 n2 n3 n4 n5 n6 n7 n8 n9]);
        if(nnd < max_n); nnd = max_n;
    end;
    %
    XI0(2*i-1) = geom(n1,1); XI0(2*i) = geom(n2,1); XI0(2*i+1) = geom(n3,1);
    YI0(2*j-1) = geom(n1,2); YI0(2*j) = geom(n4,2); YI0(2*j+1) = geom(n7,2);
end
end
    
```

Figure 5. Mesh generation of Pave-Ut and code with horizontal strain distribution.

in all directions. Assumptions are made for each layer to have linear elastic and cross-anisotropic material models. The analysis is based on axisymmetric geometry and formulation. Full bonding on all layers is assumed (Fig. 5).

Verification of Pave-Ut

For the verification of Pave-Ut, among the popular pavement analysis programs, BISAR and ALVA were chosen. The pavement parameters used in the analysis are described in Table 3. A load of 0.7 MPa was considered in a circular area with a radius of 150.8 mm.

Pave-Ut shows reliable results when compared with the other programs. However, there is a tendency to overestimate the horizontal strain and slightly underestimate the vertical strain. The

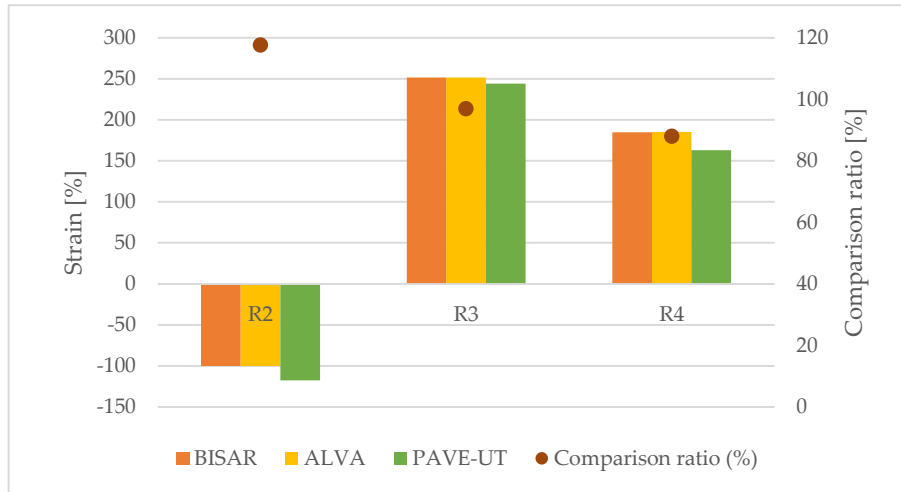


Figure 6. Comparison of structural responses; R2: horizontal strain at the bottom of the asphalt layer; R3: vertical strain at the top of the base layer; R4: vertical strain at the subgrade ($\mu\epsilon$) and comparison ratio between Pave-Ut and BISAR.

comparison ratio is marked in **Fig. 6**, which shows the ratio of the responses calculated with Pave-Ut to those calculated with BISAR.

III. RESULTS

With material properties that reflect the climatic conditions from Section 2.1, Pave-Ut carried out a structural analysis. Once the dynamic moduli of asphalt layers with specific depths were calculated, these material properties were applied to the structural analysis stage.

The basic section and physical properties for ‘K’ (heavy traffic) and ‘N’ (normal usage) roads specified in the Hungarian specifications are as follows (**Table 4**).

For the parametric study to check the influence of environmental factors on the structural responses, two cross sections are chosen: 1) simple model with simple material properties guided in Hungarian specification (**Table 4**), and 2) material properties calculated upon temperature profile on asphalt mix layer and water content change on subgrade.

Table 4. Basic pavement parameters suggested in Hungarian specification e-UT 06.03.13 [38].

Index	Stiffness [MPa]	Poisson’s ratio	Thickness [mm]
Asphalt	5000	0.35	160
Base	500	0.35	200
Subgrade	40	0.35	-

Table 5. Weather data from July.

Index	Average temp [°C]	Acc. Precipitation [mm]
July	23.56	124.7

The cross-sections that were compared with the simple cross-section from the Hungarian design code were (**Fig. 8**): 1) the material properties reflecting the temperature distribution of the asphalt layer and 2) those reflecting the properties of the subgrade considering the monthly average accumulated precipitation. Structural analysis was carried out based on the weather information in July. The weather data can be found below.

Table 5 refers to Hungary’s average temperatures in July and the monthly-accumulated precipitation data. These are the input data for the estimation of the subgrade modulus. **Fig. 7** shows the temperature distribution at points at 2, 5, and 10 cm from the surface according to the prediction model in Equation 12. With this distribution, the variation in the dynamic modulus inside the asphalt layer can be traced. These climatic conditions were applied in the calculation of the structural response.

The influence of environmental factors was evaluated and confirmed for two layers: the asphalt layer and the subgrade. As a result of confirming the results, the influence of environmental factors on road structural analysis was significant.

Table 6 shows the results of the application of load and climatic conditions in July to the pavement structure. According to the results shown in **Table 6**, the influence of the asphalt stiffness on the performance of the flexible pavement when considering temperature change is greater than that of the change in the subgrade modulus when considering water content. Critical points may occur at the interface of the asphalt layer and base layer, and a bottom-up crack is expected.

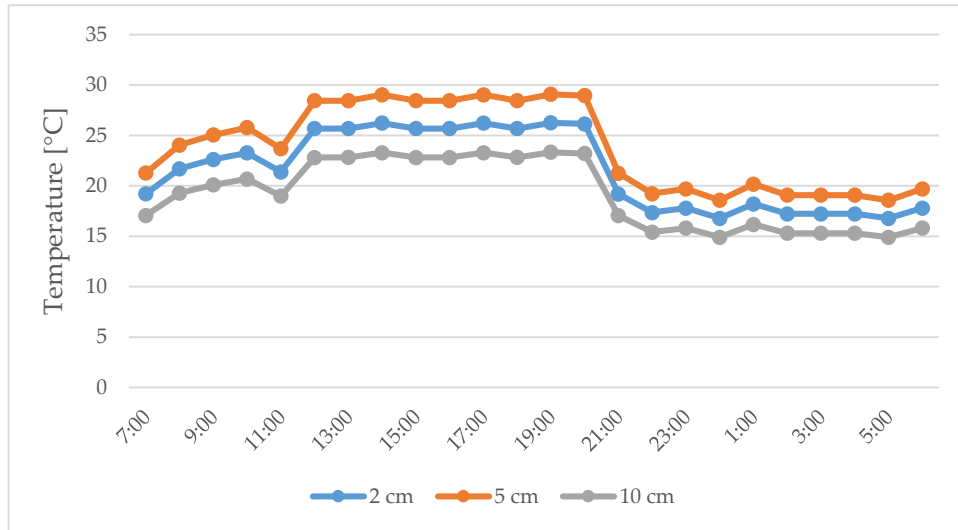


Figure 7. Estimated temperature inside of the asphalt mixture layer on July 1st according to the Hungarian estimation model (Equation 12).

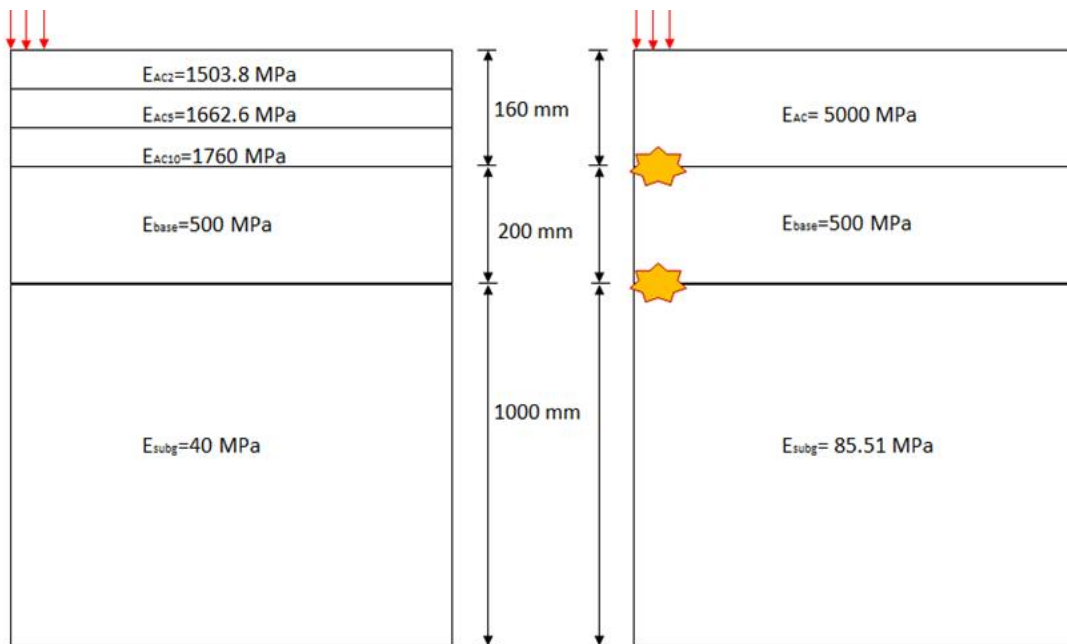


Figure 8. Material properties and cross section: (a) Variations in the predicted modulus: asphalt temperature profile considering stiffness; (b) subgrade modulus with consideration of precipitation in July. The critical points are indicated with star shapes.

Table 6. Structural responses of three cases of Figure 9.

Index	Hungarian spec	Varying Asphalt stiffness	Subgrade
Horizontal strain bottom of asphalt layer	150.7	124.4	149.5
Vertical strain at the top of base	-337.5	-131.1	-334.7
Vertical strain at the top of subgrade	-493	-213.2	-489

IV. CONCLUSION

In this study, a dynamic modulus estimation program was prepared. The Hungarian asphalt temperature estimation model and Witczak and Bari's viscosity and dynamic modulus prediction models were used. The results of this simple calculator can be used for the analysis of the aging of pavement. Here in this study research focuses on temperature and water content influence among the environmental factors that influencing pavement structure.

This study was conducted on the quantification of environmental loads for road design in Hungary. Since regional characteristics are important in the application of environmental loads to structural analysis, Hungarian weather information was actively used. The validation of the results was confirmed by comparing them with those of BISAR and ALVA, which are research codes. There was a tendency to overestimate the horizontal strain at the bottom of the asphalt layer and to slightly underestimate the vertical strain at the top of the base layer. We intend to improve this through future research.

This program has the advantage that it can be used for structural analysis by predicting physical properties by reflecting the temperature distribution inside an asphalt mixture layer.

The influences of considering the temperature profile of the asphalt layer and the variation of the water content on the subgrade noticeably exist. The strains differ from those suggested in the suggestion of a simple cross-section in the design specifications when taking the climatic conditions into account in the analysis. As a result of the confirmation of the

findings, it is clear that environmental elements have a substantial impact on road structural analysis.

ACKNOWLEDGEMENT

The authors appreciate the academic support of Dr. András Gulyás in this study.

DATA AVAILABILITY STATEMENT

The program prepared in this study can be downloaded at:

<https://tomok4737.tistory.com/category/Archive/Dynamic%20modulus%20calculator>.

Anybody who is interested in this program can access `PaveUt.exe`. The password is: 45NDQxNT.

AUTHOR CONTRIBUTIONS

S. Cho: Software, Validation, Data curation, Writing, Methodology, Investigation, Visualization.

C. Tóth: Supervision, Conceptualization, Formal analysis, Investigation, Resources, Review.

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

If the authors have ORCID identification, it may be given in this section.

S. Cho <http://orcid.org/0000-0003-2090-3218>

C. Tóth <http://orcid.org/0000-0001-5065-5177>

REFERENCES

- [1] Y. Huang, Pavement design and analysis, New York: Prentice Hall, 2004.
- [2] J. Chao and Z. Jinxi, "Prediction Model for Asphalt Pavement Temperature in High-Temperature Season in Beijing," *Advances in Civil Engineering*, pp. 1-11, 2018. <https://doi.org/10.1155/2018/1837952>
- [3] N. Afanasieva, M. Álvarez and M. J. Ortiz, "Rheological characterization of aged asphalts," *CT&F - Ciencia, Tecnología y Futuro*, 2 (3), pp. 121-135, 2002.
- [4] J. Bari and W. M. Witczak, "New Predictive Models for Viscosity and Complex Shear Modulus of Asphalt Binders," *Journal of the Transportation Research Board*, 2001, pp. 9-19, 2007. <https://doi.org/10.3141/2001-02>
- [5] H. A. Hawesah, M. Sadique, C. Harris, H. A. Nageim, K. Stopp, H. Pearl and A. Shubbar, "Improving the Temperature Sensitivity of Bitument for Emergency pavement Repair," *periodica Polytechnica Civil Engineering*, 65 (4), pp. 1190-1199, 2021. <https://doi.org/10.3311/PPci.18351>
- [6] J. Jakob and J. Tick, "Camera-based On-Road Detections for the Visually Impaired," *Acta Polytechnica Hungarica*, 17 (3), pp. 125-146, 2020. <https://doi.org/10.12700/APH.17.3.2020.3.7>
- [7] E. S. Barber, "Calculation of Maximum Pavement Temperature from Weather Reports," *Highway Research Board Bulletin*, 1957.
- [8] J. E. Dickinson, "A method for calculating the temperature gradients in asphaltic concrete pavement structures based on climatic data," *Austrian Road Research*, 8 (4), pp. 16-34, 1978.

- [9] H. B. Saas, "A numerical forecasting system for the prediction of slippery roads," *Journal of Applied Meteorology*, 36, pp. 801-817, 1997. [https://doi.org/10.1175/1520-0450\(1997\)036<0801:ANFSFT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1997)036<0801:ANFSFT>2.0.CO;2)
- [10] C. Louis-Philippe and Y. Delage, "METRo: A New Model for Road-Condition Forecasting in Canada," *Journal of Applied Meteorology and Climatology*, 40 (11), pp. 2026-2037, 2001. [https://doi.org/10.1175/1520-0450\(2001\)040<2026:MANMFR>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<2026:MANMFR>2.0.CO;2)
- [11] J. A. Deacon, J. S. Coplantz, A. A. Tayebali and C. L. Monismith, "Temperature Considerations in Asphalt Aggregate Mixtures Analysis and Design," *Transportation Research Record*, 1454, pp. 97-112, 1994. <http://onlinepubs.trb.org/Onlinepubs/trr/1994/1454/1454-013.pdf>
- [12] B. J. Dempsy, "Characterizing Temperature Effects for Pavement Analysis and Design," *Transportation Research Record*, 1987. <https://onlinepubs.trb.org/Onlinepubs/trr/1987/1121/1121-003.pdf>
- [13] M. Solaimanian and T. W. Kennedy, "Predicting maximum pavement surface temperature using maximum air temperature and hourly solar radiation," *Transportation Research Record*, Washington, 1993. <http://onlinepubs.trb.org/Onlinepubs/trr/1993/1417/1417-001.pdf>
- [14] C. Marshall, R. Meier and WelchMichael, "Seasonal Temperature Effects on Flexible Pavements in Tennessee," *Transportation Research Record*, 1764, pp. 89-96, 2001. <https://doi.org/10.3141/1764-10>
- [15] W. M. Witzak, "Design of full-depth asphalt airfield pavements," *Third International Conference on the Structural Design of Asphalt Pavements*, London, 1972.
- [16] M. J. Duncan, C. L. Monismith and E. L. Wilson, "Finite Element Analyses of Pavements," *Highway Research Record*, 228, pp. 18-33, 1968.
- [17] R. G. Hicks and C. L. Monismith, "Factors influencing the resilient response of granular materials," *Highway research record*, 345, 1971.
- [18] J. Uzan, "Characterization of Granular Material," *Transportation Research Record*, Washington D.C., 1985.
- [19] W. M. Witzak and J. Uzan, "The universal airport pavement design system," *University of Maryland, Maryland*, 1988.
- [20] B. H. Seed, G. F. Mitry and L. C. Monismith, "Factors influencing the resilient deformations of untreated aggregate base in two layer pavements subjected to repeated loading," *Highway Research Record* 190, 1967, pp. 19-57.
- [21] S. Helwany, J. Dyer and J. Leidy, "Finite-Element Analysis of Flexible Pavements," *Journal of transportation engineering*, 124 (5), pp. 491-499, 1998.
- [22] A. Skar and S. Andersen, "ALVA: An adaptive MATLAB package for layered viscoelastic analysis," *Journal of Open-Source Software*, 5 (55), pp. 1-3, 2020. <https://doi.org/10.21105/joss.02548>
- [23] H. Huang, M. Moaveni, I. I. Qamhia, E. Tutumluer and S. J. Tingle, "Advanced Analytical Tool for Flexible Pavement Design and Evaluation," *Airfield and Highway Pavements*, pp. 61-69, 2019. <https://doi.org/10.1061/9780784482452.007>
- [24] L. Raad and L. Figueroa, "Load response of transportation support systems," *Transportation Engineering Journal of ASCE*, 16 (1), pp. 111-128, 1980.
- [25] L. Gáspár, "Lifetime Engineering for Roads," *Acta Technica Jaurinensis*, pp. 37-46, 2008.
- [26] B. Eller, M. Movahedi Rad and S. Fischer, "Laboratory Tests and FE Modeling of the Concrete Canvas, for Infrastructure Applications," *Acta Polytechnica Hungarica*, 19 (3), pp. 9-20, 2022. <https://doi.org/10.12700/APH.19.3.2022.3.2>
- [27] W. R. Gloyne, "The diurnal variation of global radiation on a Horizontal surface with special reference to Aberdeen," *Meteorological Magazine*, 101, pp. 44-51, 1972.
- [28] E. Kreyszig, H. Kreyszig and E. J. Norminton, *Advanced engineering mathematics*, John Wiley & Sons, 2011.
- [29] NCHRP, "LTPP Computed Parameter: Dynamic modulus," *US Department of Transportation Federal Highway Administration*, 2011.
- [30] AASHTO, *Mechanistic-Empirical Pavement Design Guide*, Washington DC: American Association of State Highway and Transportation Officials, 2008.
- [31] W. M. Witzak, D. Andrei and N. W. Houston, "Resilient Modulus as Function of Soil Moisture-Summary of Predictive Models," *NCHRP 1-37A*, Arizona, 2000.
- [32] E. C. Zapata, D. Andrei, W. M. Witzak and N. W. Houston, "Incorporation of Environmental Effects in Pavement Design," *Road Materials and Pavement Design*, 8(4), pp. 667-693, 2006. <https://doi.org/10.1080/14680629.2007.9690094>
- [33] MOLIT, *Korean Pavement Design Guide Phase 4.*, Sejong city: Ministry of Land, Transport, and Maritime Affairs, 2011.
- [34] C. E. Cary and C. E. Zapata, "Enhanced Model for Resilient Response of Soils Resulting from Seasonal Changes as Implemented in Mechanistic-Empirical Pavement Design Guide," *Transportation Research Board of the National Academies*, Washington D.C., 2010.

- [35] G. Nikishkov, Programming Finite Elements in Java, Aizu: Springer, 2010.
- [36] I. Bojtár and G. Zsolt, The Finite Element Method-The Basis, Budapest: Budapest University of Technology and Economics, 2011.
- [37] M. Kim, Three-dimensional finite element analysis of flexible pavements considering nonlinear pavement foundation behavior, Illinois: University of Illinois at Urbana-Champaign, 2007.
- [38] MOLIT, Mechanistic-empirical pavement design specification, Seoul: Ministry of Land, Infrastructure and Transport, Korea, 2015.



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.

Sensitivity of Geometric Parameters in the Sustainability Development of Continuous Welded Rail

Katarzyna Dybel^{1,*}, Arkadiusz Kampczyk¹

¹ AGH University of Science and Technology,
Faculty of Geo-Data Science, Geodesy, and Environmental Engineering,
Department of Engineering Surveying and Civil Engineering,
al. A. Mickiewicza 30, 30-059 Krakow, Poland

*e-mail: kdybel@agh.edu.pl, katarzynadybel@gmail.com

Submitted: 29/04/2022 Accepted: 27/06/2022 Published online: 15/07/2022

Abstract: Continuous Welded Rails (CWR) are a key infrastructure element in the safety and efficiency of rail transportation. Their correct exploitation (operational) requires surveying and diagnostic monitoring based not only on the results of rail displacement measurements, but also on the geometric parameters of the track in the horizontal (H) and vertical (V) planes. Many researchers have proposed different approaches for surveying and diagnostic monitoring of CWR. However, they do not refer to the determination of railway track defectiveness (parametric defects, track defectiveness) respectively on straight and curvilinear segments. Research topics involving CWR constitute a continuous openness to research with particular application of synergy effects in the optimization of monitoring of CWR geometry shaped by exploitation processes. In this study, based on real measurement data of six geometric parameters (H: track gauge, gradient of track gauge, horizontal irregularities and V: cant, twist, vertical irregularities), the most sensitive parameters in sustainable development CWR are defined. The research answered that the most sensitive parameters in the sustainability development of CWR belong in the range of the plane H: gradient of track gauge and horizontal irregularities, and in the plane V: vertical irregularities. These escalate especially on curvilinear sections, requiring more significant maintenance capacity. Due to the growing importance of rail transportation as a sustainable, environmentally friendly, and mass transit mode, the research results provide a basis for life cycle management of CWR.

Keywords: *Continuous Welded Rail; CWR; Geometrical parameters; Railway track defectiveness; Surveying*

I. INTRODUCTION

Developments in infrastructure technology provide new ways to solve existing problems in an efficient and cost-effective manner [1], [2], [3]. Investments in rail transport infrastructure often in the technical field are related to replacement of conventional railway tracks (track) with Continuous Welded Rails (CWR) in order to increase passenger comfort, reduce train running time and thus make rail more competitive in the transport sector. Research on ensuring the stability of CWR and improving its condition monitoring systems is an active area of research that is of great importance and constantly evolving [4], [5], [6]. Sabato and Niezrecki in [7] note that it is important to monitoring the construction condition (Structural Health Monitoring, SHM) of aging railway tracks.

Mrówczyńska et al. in [8] aptly state that regardless of the results and their interpretation, we should remember that in the case of geodetic measurements and continuous or periodic geodetic monitoring, the choice of the measurement method and methods of data processing is determined by the nature of the object and specific terrain conditions. In turn, Shvets in [9] notes that experimental studies, while sufficiently reliable, are resource-intensive, time-consuming and cannot cover all situations that may occur during exploitation. Theoretical and experimental research are complementary and should be conducted together.

The major feature of CWR design is the presence of temperature stresses in the rails, which affect the stability parameters during its operation (exploitation). Violations committed during the construction of a CWR, as well as in the process of

current maintenance, together with the impact of train load and natural and climatic factors, create conditions for the growth of temperature stresses in rails and changes in the temperature regime of their operation [10]. The combination of these factors can finally lead to deformation or damage to the railway track structure, consequently risking the safety of operations. An important role in preventing damage propagation in CWR is the regular monitoring of the geometry and temperature condition of the rails [11], [12]. Kukulski et al. in [13] presented an effective analytical method for diagnosing the condition of CWR based on experimental measurements. Supporting the decision process in the area of track repair or maintenance. In addition, in [13] they recognise that in the case of a CWR, the costs of its maintenance are about 25% lower than those of a classic track. There is also significantly less wear and tear on vehicles and traction energy consumption, with better ride smoothness and less noise. The interest in the issue of CWR is connected, among others, with the change in operating conditions of modern railway by increasing the speed of passenger trains and the permissible axle loads on the rails. The CWR, in comparison with the classic track, provides better ride smoothness together with a reduction of the acoustic wave emission associated with the passage of the train [13].

Most railways are described in terms of piece-wise curves [14]. The application of appropriate geometry is particularly important for tracks in operations, modernization, revitalization and during repair work. Directly reflected in the comfort and safety of transportation and in the maintenance needs and sustainable cycle of their life management. Hasan in [15] derived formula to determine the threshold value: the temperature limits at which hot-weather or cold-weather patrolling is to be enforced if the radius happens to be sharper than the threshold radius. Doyle and Thomet in [16] note that passenger comfort is an important constraint on high-speed operation in curves and transitions. A detailed analysis of methods of track stability estimation for small radius curves is presented in papers [15], [17]. Conditions for the safe exploitation of CWR require monitoring their geometric parameters, especially in plane:

1. horizontal (H):

- track gauge,
- gradient of track gauge (track gauge gradient),
- irregularities of track rails in the horizontal plane (horizontal irregularities),

2. vertical (V):

- cant (superelevation),
- twist,
- irregularities of track rails in the vertical plane (vertical irregularities).

To the factors threats to the stability of the CWR include, among others:

- temperature spikes,
- longitudinal displacements of the rails (creeping of the rails, pl. *pełzanie toków szynowych*),
- incorrect technical condition,
- operational impacts.

Failure to comply with normative construction conditions and maintenance of the CWR combined with the effects of operational influences and weather conditions can lead to disorders balance of forces inside the railway track structure. Increasing the risk of deformation. A key role in providing CWR stability and driving comfort of rolling stock is played by surveying and diagnostic monitoring. The sustainability development of the research topic undertaken is a development that meets the needs of scientists, researchers, end users, industry and branch-users, decision makers from many countries and professional backgrounds without compromising the ability of future generations, and especially in correspondence of ensuring the CWR needs that occur. The idea of sensitivity of geometric parameters in the sustainability development of Continuous Welded Rail is to support sustainable decision making for rail infrastructure. One of the main challenges facing the world today in terms of the development of transport components is sustainability development. The most current challenge to be met for a fully functional rail transport is seamlessly connected infrastructure components in a sustainable development system. The main pillar of rail transport infrastructure is the railway track, which requires measures to ensure safe exploitation.

The main objective of this research is to define the most sensitive geometric parameters in the H and V plane escalating especially in straight and curvilinear segments in the sustainability development of CWR. The research included analysis and evaluation defectiveness of geometric parameters in straight and curvilinear segments (transition curves/spirals, circular curves, compound curve). Based on real measurement data of six geometric parameters (H: track gauge, gradient of track gauge, horizontal irregularities and V: cant, twist, vertical irregularities).

Railway track defectiveness (parametric defects, track defectiveness) is a relative measure of railway track condition. It depends on the maximum speed of the trains V_{max} . Its formula represents the ratio of the railway track length – on which the permissible deviations are exceeded – to the total length considered. Railway track defectiveness refers to the individual geometric parameters in the H and V planes. Railway track defectiveness is grounded in regulation Id-14 (D-75) [18]. It define the defectiveness of each parameter on the baseline segment being evaluated as the ratio of the sum of the lengths of the segments where allowable

deviations are exceeded to the total length of that segment. For each measured railway track parameter, the defectiveness is (1) [18]:

$$W = \frac{n_p}{n} \quad (1)$$

where:

n_p – number of signal samples exceeding the permissible deviations in the analyzed segment,

n – number of signal samples on the analyzed segment.

Research was conducted on three real objects – railway lines, including five research objects – railway tracks characterized by different geometric features in plan and profile as well as operational elements of transportation engineering (**Table 1 - 3**):
 Object no. R_{143_1_1} and R_{143_1_2} Track no. 1
 Object no. R_{143_1_2} Track no. 2
 Object no. R_{144_1} Track no. 1

Object no. R_{161_1}

Track no. 1

Object no. R_{161_2}

Track no. 2

Table 1. Technical and exploitation specification of research objects

<i>Technical and exploitation parameters of research objects</i>	
<i>Name</i>	<i>Values</i>
Exploitation load T	$10 \leq T < 25$
Speed of passenger trains V_{max}	$80 < V_{max} \leq 120$
Speed of freight trains V_{ft}	$60 < V_{ft} \leq 80$
Permissible axle loads P	$210 \leq P < 221$

where:

T – in [Tg/per year]

V_{max} – in [km/h]

V_{ft} – in [km/h]

P – in [kN]

Table 2. Characteristics of selected railway track - permanent way elements of research objects

<i>Research object</i>	<i>Rail type</i>	<i>Sleeper type</i>	<i>Type of rail fastening</i>	<i>Historic inauguration</i>	<i>Type of refurbishment</i>
Object no. R _{143_1_1} and R _{143_1_2}	60E1	wooden (along the 'protective section' length)	Skl (along the 'protective section' length)	1884 ¹⁾ 1890 ³⁾	RM – 2014
		PS94	SB		
Object no. R _{143_2}	60E1	wooden (along the 'protective section' length)	Skl (along the 'protective section' length)	1896 ¹⁾ 1890 ³⁾	RM – 2015
		PS94	SB		
Object no. R _{144_1}	UIC60	wooden	K	1857 ²⁾ 1964 ³⁾	RM – 1980 RR – 2019
Object no. R _{161_1}	60E1	PS83	SB	1874 ³⁾	RM – 2020
Object no. R _{161_2}	60E1	PS83	SB	1874 ³⁾	RM – 2020

where:

RM – refurbishment – master repair

RR – refurbishment – running repair

¹⁾ – opening for exploitation (put into service)

²⁾ – technical acceptance

³⁾ – on the basis of the railway infrastructure quantity evidence (records) – fixed asset

The identification of real objects preserves their anonymity so that no direct identification can be made for a given building object.

Results at all research objects include one measurement period – 2021. The authors verified hypotheses related to the sensitivity of geometric parameters in the sustainability development of Continuous Welded Rail, i.e:

- the possibility to inadequate maintenance of CWR may affect the track defectiveness,
- the possibility to indicate CWR segments most vulnerable to exceeding permissible deviations,
- the possibility to select the most sensitive geometrical parameters of CWR.

Table 3. Characteristics of actual and research objects

Characteristics of a curvilinear and straight segment	Kilometer [km]	Geometry of the curvilinear segment	Characteristics of the research object					
			Speed [km/h]	Category of railway line	Type of railway line / movement			
Research object: Object no. R143_1,1 and R143_1,2								
Curvilinear segment:								
Transition curve	1+006.22 ÷ 1+036.22	L ₁ = 30.00 m D ₁ = 282.80 m R ₁ = 1720.00 m	V _{max} = 120 V _{ft} = 100	prime	double-track, single direction			
Compound curve	1+036.22 ÷ 1+357.02	D ₂ = 38.00 m R ₂ = 2050.00 m						
Transition curve	1+357.02 ÷ 1+417.02	L ₂ = 60.00 m C _{ant t.} = 30 mm						
Straight segment:								
Straight	1+417.02 ÷ 2+517.02	D ₃ = 1100.00 m						
Research object: Object no. R143_2								
Curvilinear segment:								
Transition curve	0+986.04 ÷ 1+056.04	L ₁ = 70.00 m D ₁ = 248.40 m R ₁ = 1740.00 m	V _{max} = 120 V _{ft} = 100	prime	double-track, single direction			
Circular curve	1+056.04 ÷ 1+304.44	L ₂ = 150.00 m C _{ant t.} = 30 mm						
Transition curve	1+304.44 ÷ 1+454.44							
Straight segment:								
Straight	1+454.44 ÷ 2+554.44	D ₂ = 1100.00 m						
Research object: Object no. R144_1								
Curvilinear segment:								
Transition curve	14+860.00 ÷ 14+980.00	L ₁ = 120.00 m D ₁ = 660.00 m R ₁ = 1090.00 m	V _{max} = 70 V _{ft} = 70	prime	single-track, double direction			
Circular curve	14+980.00 ÷ 15+640.00	L ₂ = 120.00 m C _{ant t.} = 40 mm						
Transition curve	15+640.00 ÷ 15+760.00							
Straight segment:								
Straight	15+760.00 ÷ 16+860.00	D ₂ = 1100.00 m						
Research object: Object no. R161_1								
Straight segment:								
Straight	4+411.24 ÷ 4+611.24	D ₁ = 200.00 m	V _{max} = 70 V _{ft} = 70	prime	double-track, single direction			
Curvilinear segment:								
Transition curve	4+611.24 ÷ 4+651.24	L ₁ = 40.00 m D ₁ = 55.78 m R ₁ = 3300.00 m						
Circular curve	4+651.24 ÷ 4+707.02	L ₂ = 40.00 m C _{ant t.} = 20 mm						
Transition curve	4+707.02 ÷ 4+747.02							
Straight segment:								
Straight	4+747.02 ÷ 4+847.02	D ₂ = 100.00 m						
Straight	9+500.00 ÷ 10+300.00	D ₃ = 800.00 m						
Research object: Object no. R161_2								
Straight segment:								
Straight	4+408.69 ÷ 4+608.69	D ₁ = 200.00 m	V _{max} = 70 V _{ft} = 70	prime	double-track, single direction			
Curvilinear segment:								
Transition curve	4+608.69 ÷ 4+648.69	L ₁ = 40.00 m D ₁ = 57.69 m R ₁ = 3300.00 m						
Circular curve	4+648.69 ÷ 4+706.38	L ₂ = 40.00 m C _{ant t.} = 20 mm						
Transition curve	4+706.38 ÷ 4+746.38							
Straight segment:								
Straight	4+746.38 ÷ 4+846.38	D ₂ = 100.00 m						
Straight	9+500.00 ÷ 10+300.00	D ₃ = 800.00 m						

where: L_i – length of the transition curve, D_i – straight or curve length, R_i – radius, V_{max} – speed of passenger trains, V_{ft} – speed of freight trains, C_{ant t.} – theoretical cant value

The research has provided the answer that the inadequate maintenance of the shape of the gravel prism layers in CWR are reflected in the railway track defectiveness, especially in the gradient of track gauge. To the most sensitive parameters in the sustainability of CWR belong in the range of the plane H: gradient of track gauge and horizontal irregularities, but in the plane V: vertical irregularities. These escalate especially on curvilinear segments, requiring more significant maintenance capacity. Completed research provides a source of knowledge and support for optimization of Continuous Welded Rails geometry monitoring.

II. METHODS AND APPLIED MATERIALS

The search for innovations in rail transport results in better use of existing potential and shapes new development perspectives. Optimization in maintenance systems CWR contributes to the efficient and sustainable development of transport infrastructure, while creating new opportunities for transport and logistics operators. In order to make measurements and research on the sensitivity of geometric parameters in the sustainability development of CWR, a direct method was used – mobile measuring equipment using electronic self-recording track gauge TEC-1435 N2 (**Fig. 1**, **Fig. 2**) and for measurements of the displacement of CWR rails used a measurement method called the fixed point method.



Figure 1. Electronic self-recording track gauge TEC-1435 N2



Figure 2. Electronic self-recording recorder of track gauge TEC-1435 N2

The construction of the track gauge is three-point, and direct measurements are made without load. The performed measurements provided the parameters of the track geometry in the plane H: track gauge, gradient of track gauge, horizontal irregularities and in the plane V: cant (position of the track in the cross-section), twist, vertical irregularities.

In terms of the measurement work carried out, the research also included monitoring of such elements as: lack of sleepers in a given kilometer of the railway line, weed, breaks – broken rail and local depressions of the running surface (squat), lack of screws, prism condition, etc., and locations of features/events around existing ones, e.g., railway level crossing, bridges, overpasses, tunnels, hectometric points, etc.

The scientific and research work on the topic of sensitivity of geometric parameters in the sustainability development of Continuous Welded Rail was carried out in correspondence with legal regulations:

- Instruction for Measuring, Researching and Assessing Track Condition Id-14 (D-75) [18],
- Technical Conditions for Maintaining Track Surface on Railway Lines Id-1 (D-1) [19],
- Instruction for railway superstructure diagnosis Id-8 [20].

Direct measurements in accordance with regulations Id-14 (D-75) [18] i Id-1 (D-1) [19] are made:

- twice a year (spring and autumn) – for all railway tracks of each category. Direct measurements are not performed when a detour with a measuring vehicle, e.g. a measuring trolley (Measure track motor car, Measuring motor car, Track recording cars, Real time systems for geometry cars), is planned,
- once every six months – measurement of railway tracks in curves with radius $350 \div 500$ meters on railway lines of all categories,
- once every four months – measurement of railway tracks in curves with radius less than 350 meters on railway lines of all categories.

Measurements of the displacement of CWR rails as places susceptible to creeping of the rail were carried out using a measurement method called the fixed point method [19], [20]. The purpose of this method was to determine the value of the creeping of the rails in relation to fixed points. The measurements were conducted in all railway tracks of the research objects – once a year before the period of increased temperatures.

Results for all objects include one measurement period – spring 2021 r. using the direct method and the fixed point method (**Fig. 3**):

Object no. R_{143_1_1} and R_{143_1_2}
Object no. R_{143_1_2}

Track no. 1
Track no. 2

Object no. R_{144_1} Track no. 1
 Object no. R_{161_1} Track no. 1
 Object no. R_{161_2} Track no. 2

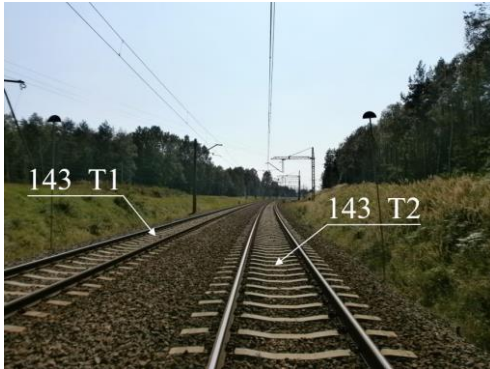


Figure 3. Actual and research object: Object no. R_{143_1_1} and R_{143_1_2} and Object no. R_{143_1_2}

III. RESULTS AND DISCUSSION

Conducting research on the sensitivity of geometric parameters in CWR sustainability development is applicable to rail transportation engineering, while providing effective support for optimizing the monitoring of such structures. It significantly supports the process of correct analysis and assessment of the technical condition of rail transport infrastructure. Implies correct diagnosis and correctness of actions taken to prevent the development of further damage threatening the safe operation of the construction.

The research conducted in this publication, while combining scientific and practical knowledge and correlating with the results of research Shvets in [9], that “theoretical and experimental research are complementary and should be conducted together” are part of the application of synergy effects in optimizing the monitoring of CWR geometry shaped by operational processes.

The reason for the occurrence of a high level of defectiveness of the railway track of particular geometrical parameters may be, for example, the weakening of the construction caused by the slack in the rail fastenings, slight resistance of the ballast, utility damage, lack of ballast – insufficient maintenance of the shape of the ballast prism of the railway superstructure, etc. Research in this direction leads to the identification of CWR segments most exposed to the occurrence of permissible deviation exceedances and to the selection of the most sensitive geometric parameters in the sustainability development of Continuous Welded Rail.

CWR stability in sustainability development requires not only surveying and diagnostic monitoring of longitudinal rail displacements (creep of the rail), but also geometrical parameters in the horizontal and vertical plane. Longitudinal displacements of the rails have an influence on the

stability of the CWR. This is confirmed by Towpik in [21], stating that longitudinal displacements of the rails and local horizontal and vertical deformations of the track increasing over time can lead to local accumulation of stresses causing deformation of the track frame leading in extreme cases to loss of stability by the CWR.

The research topic covering the sensitivity of geometric parameters in the sustainability development of Continuous Welded Rail, then its objectives, including specific hypotheses and its relevance are in demand not only in the scientific and research community, but also in the branch. The idea of the Authors is to ensure that the research results are communicated and understood also by researchers and related industry professionals not involved in this publication topic. Therefore, a detailed expression of them in % has been made to enable others to replicate and use the published results. As a result of the analysis and evaluation of the acquired geometric data in the CWR, a histogram of permissible deviation exceedances was developed, distinguishing straight and curvilinear segments (curves and transition curves) of all research objects (Fig. 4).

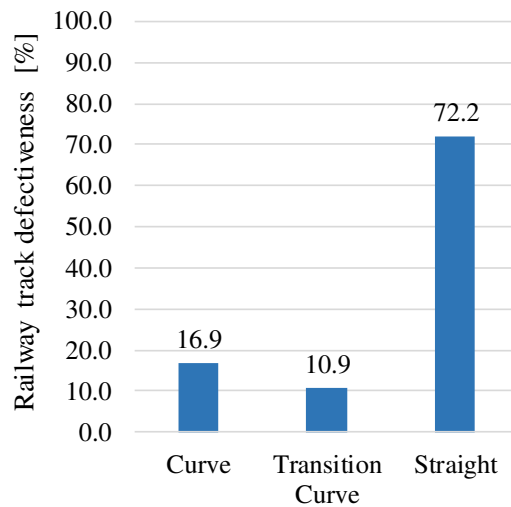


Figure 4. Histogram of permissible deviation exceedances in the measured data samples all research objects

Fig. 4 shows a significant spike in the level of railway track defectiveness on straight sections up to 72.2 %. This spike is determined by the number of parametric defects in the gradient of track gauge in Object no. R_{143_1_1} and R_{143_1_2}. Inadequate maintenance of gravel prism layer formation – lack of ballast in CWR in Object no. R_{143_1_1} and R_{143_1_2} is directly reflected in the parametric defects of the whole set of exceeded permissible deviations in all straight segments of the research. Numerous defects in the gravel, up to a value of 23 cm relative to the top surface of the sleeper, contribute to the disturbed geometry of this research object (Fig. 5). At the same

time having a direct reflection in straight and curvilinear segments of the main aim of the research – which is to define the most sensitive geometric parameters in the H and V plane escalating especially in straight and curvilinear segments in the sustainability of CWR. Object no. R_{143_1_1} and R_{143_1_2} disrupted the insights of the research as a whole. Inadequate maintenance of CWR through proper gravel prism profiling has an impact on the track defectiveness. Thus confirming the hypothesis of the possibility that insufficient maintenance of CWR on the track defectiveness.

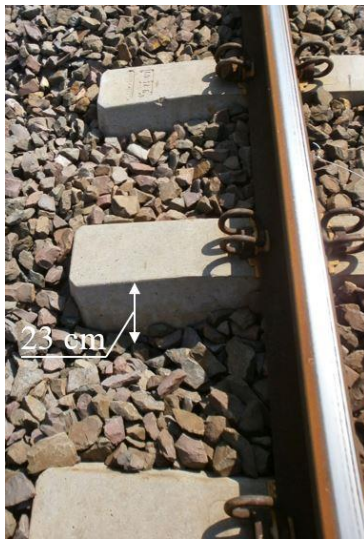


Figure 5. Loss of ballast in Object no. R_{143_1_1} and R_{143_1_2} – straight segment

Therefore, in continuing the analysis and evaluation of geometric parameters in CWR, this straight segment was removed – obtaining a histogram of permissible deviation exceedances in the measured data sample without a straight section of Object no. R_{143_1_1} and R_{143_1_2} (**Fig. 6**). As a result of removing the disturbed straight segment from research Object no. R_{143_1_1} and R_{143_1_2}, obtained insights into the totality of all research objects. From which it is concluded that the number of parametric defects in CWR escalates significantly in curvilinear segments, embracing in the curves 16.9 % and in the transition curves 10.9 %, giving a total of 27.8 %. These segments are more susceptible to defects than the straight segments 13.9 %. At the same time, it should be taken into account that the histogram contained in the fig. 6 includes all research objects, excluding only straight segment of Object no. R_{143_1_1} and R_{143_1_2}. So, at this stage of the research, it represents a certain deficiency in the proportionality of the analogy of straight versus curvilinear segments.

Consequently, in order to properly define and identify the segments CWR the most exposed to the

occurrence of exceedances of permissible deviations were completely eliminated Object no. R_{143_1_1} and R_{143_1_2} from the overall analysis and evaluation of the sensitivity of geometric parameters in the sustainability development of CWR. De facto, in the final stage of the research, conducting analysis and evaluation on three real objects – railway lines, covering four research objects – railway tracks. The result obtained shows **Fig. 7**, including histogram of permissible deviation exceedances in the measured data sample without Object no. R_{143_1_1} and R_{143_1_2}.

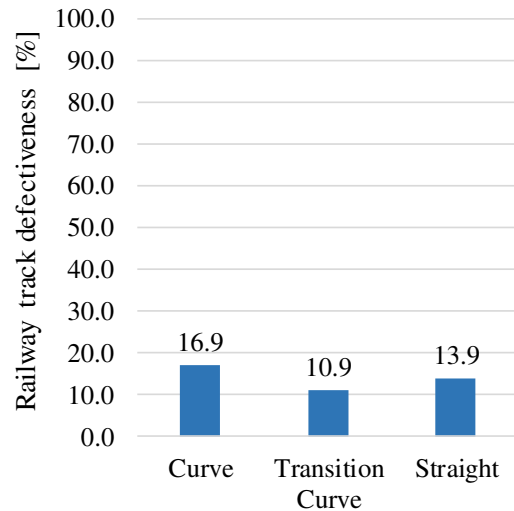


Figure 6. Histogram of permissible deviation exceedances in the measured data sample without a straight segment of Object no. R_{143_1_1} and R_{143_1_2}

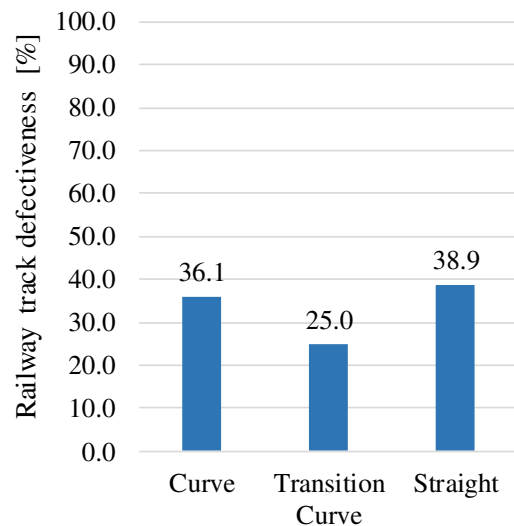


Figure 7. Histogram of permissible deviation exceedances in the measured data sample without Object no. R_{143_1_1} and R_{143_1_2}

The histogram in **Fig. 7** shows in the whole research on the verification of the hypothesis of the possibility of indicating segments CWR most susceptible to the occurrence of exceedances of the

permissible deviations, it is unambiguously that for CWR geometric parameters are more exposed on curvilinear segments than on straight segments. Track defectiveness in segments:

- curvilinear is 61.1 % (of which on curves 36.1 %, on transition curves 25.0 %),
- straight is 38.9 %.

Both figure 6 and figure 7 confirm that the highest level of defectiveness in CWR geometric parameters occurs in curvilinear segments. Conducting verification of the hypothesis regarding the possibility of selecting the most sensitive geometric parameters of CWR the results were interpreted as the percentage distribution of individual parametric defects for all research objects (Fig. 8).

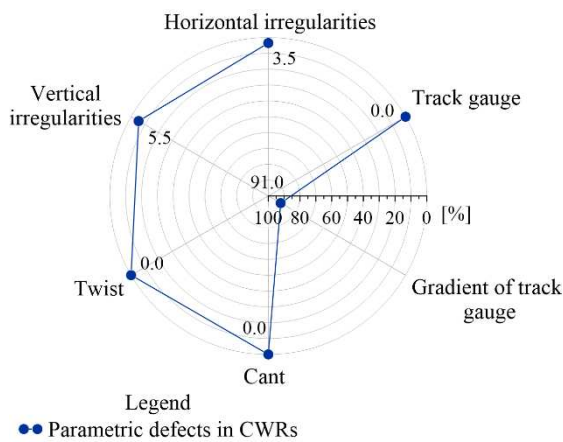


Figure 8. Percentage distribution of individual parametric defects

Conditions for safe CWR operation require monitoring of six geometric parameters. In the plane H: track gauge, gradient of track gauge, irregularities of track rails in the horizontal plane. However, in the plane V: cant, twist, irregularities of track rails in the vertical plane. Of the six geometric parameters analysed and evaluated, the most vulnerable to sensitivity escalation is:

- gradient of track gauge 91.0 % ,
- irregularities of track rails in the vertical plane 5.5 % ,
- irregularities of track rails in the horizontal plane 3.5 % .

The research answered that to the most sensitive parameters in sustainability CWR include in the plane H: gradient of track gauge and irregularities of track rails in the horizontal plane, while in the plane V: irregularities of track rails in the vertical plane.

It was also confirmed that most parametric defects exist on curves with smaller radiuses (Fig. 9):

- Object no. R_{143_1_1} and R_{143_1_2}:
R_{143_1_1} = 1720.00 m curve 23.5 %
- Object no. R_{143_1_2}:

- R_{143_2} = 1740.00 m curve 29.4 %
- Object no. R_{144_1}:
R_{144_1} = 1090.00 m curve 47.1 %

which indicates that the more complicated the track geometry, the more the probability of defects in the geometry parameters of CWR is higher. It is also rationalized by the type of line and transport engineering. Object no. R_{143_1_1} and R_{143_1_2} and Object no. R_{143_1_2} is a double-track line carrying single direction movement in each track, but both track no. 1 and 2 carry passenger and freight train movements (Table 3). In contrast Object no. R_{144_1} is a single-track line (in the analyzed kilometrage) with two-way movement, mainly freight (Table 3).

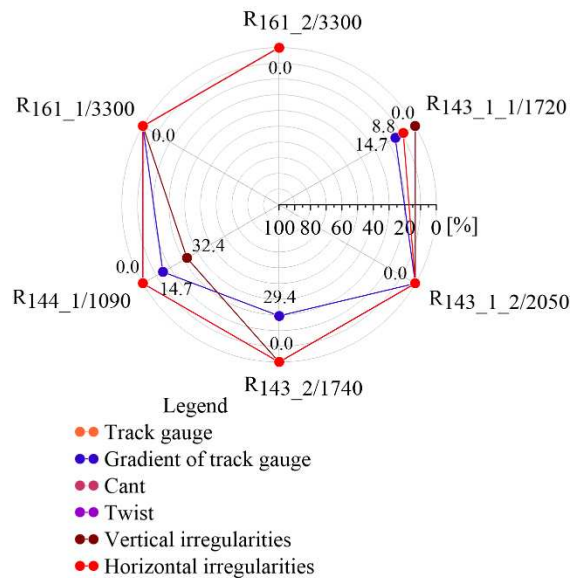


Figure 9. Percentage distribution of parametric defectivity on a given curve radius, where: R_{xxx_x/xxxx} – identification of the research object with information on the value of the curve radius (for example: R_{144_1/1090} – radius of Object no. R_{144_1} in railway track no. 1, value 1090 m)

Analogously, the research indicated that the highest number of parametric defects occurs on curvilinear segments (curves and transition curves) (Fig. 10):

- Object no. R_{143_1_1} and R_{143_1_2} 21.5 %
- Object no. R_{143_2} 35.7 %
- Object no. R_{144_1} 42.8 %

Uren and Price in [22] state that a transition curve differs from a circular curve in that its radius is constantly changing. As may be expected, such curves involve more complex formulae than curves of constant radius and their design can be complicated. Basak and Nowak in [23] dealing with dynamics of railway vehicles movement on transition curves conclude that the transition curve is a very important element of any road, including a

railway road, although it is almost invisible to traffic participants.

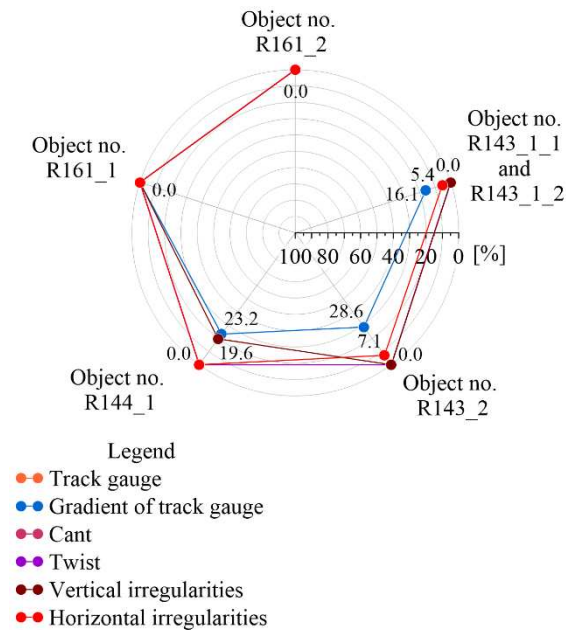


Figure 10. Percentage distribution of parametric defects on curvilinear objects

Track layout geometry (TLG), is spatial in nature, comprising [24]:

- the longitudinal horizontal plane (in plan), where straight sections, curvilinear sections, among others, are to be distinguished.
- the vertical longitudinal plane (in profile), where the following should be distinguished railway track segments with the same gradient – slope, vertical circular curves: concave or convex.
- the plane in which, due to railway practice, the system of railway track geometrical parameters comprising the vertical and horizontal railway track position is used: vertical parameters V and horizontal parameters H.

The geometrical layout of the railway track is a set of features that shape the position of the railway track in the vertical and horizontal plane, including straight and curvilinear sections. TLG at both stations and open line is dependent among others from train speeds, kinematic volumes and land availability. Land availability has an influence on the type of curvilinear sections used. At the same time, TLG has a decisive influence on the exploitation parameters of railway lines, especially maximum speed of trains, energy consumption, movement resistances [25], then on the life cycle of individual facilities, exploitation and maintenance costs of the railway track – permanent way (superstructure), and finally on the smooth running (smoothness of driving). The scientific and research work carried out, combined

with practice and in-depth discussion, has led to further conclusions:

1. Identification of the most sensitivity of geometric parameters in the sustainability development of Continuous Welded Rail

Conditions for the safe exploitation of CWR at each monitored object included the study of six geometrical parameters. Showing that the most sensitive parameters in the sustainability development of CWR include in the H-plane range: gradient of track gauge and irregularities of track rails in the horizontal plane, while in the V-plane range: irregularities of track rails in the vertical plane. These parameters are characterised by one essential element. They are continuous parameters. Parameters such as track gauge and cant do not belong to them - as they are point parameters, the identification of which takes place in a single cross-section of the railway track. In contrast, the most sensitive parameters:

- irregularities of track rail in the vertical plane,
 - irregularities of track rail in the horizontal plane,
 - gradient of track gauge,
- have a common denominator, which is the continuity of their identification basis.

Vertical irregularities, in practice called ‘holes’, are identified on a measuring base equal to 10.0 m in length. Similarly, the measurement base for identifying horizontal irregularities is also 10.0 m. In turn, the gradient of track gauge shall be defined on the basis of a measurement base of 1.0 m.

2. Interaction of defects in continuous parameters with infrastructure elements and its geometry

The railway superstructure is a construction designed to transfer to the ground the stationary and moving loads associated with the movement of railway vehicles, includes the railway track, which consists of two rails laid at a set distance, being at the same time the basic load-bearing system of the railway superstructure. In the studies conducted on the sensitivity of geometric parameters in the sustainability development, all railway tracks are Continuous Welded Rail. Railway track with welded rails with lengths of 180 m and more is CWR (Table 3) [19]. The components of the railway superstructure, especially rail type, sleeper type, type of rail fastening, ballast (Table 2) represent a particular interaction in the sustainability development of Continuous Welded Rail. The railway superstructure elements (including their constituent segments) and the location of the continuous parameter defect - argue the presence of the parameter defect:

- the gradient of track gauge, especially in linear structures with a structural assembly made of wooden sleepers, especially with K-type rail fastenings system. The study revealed the defects

and escalation of this parameter also in the structural assembly assembled with wooden sleepers and Sk1-type rail fastenings system - applied to CWR protection spans (protective section). The defectiveness of this parameter is also present in the areas of peak values of CWR creeping of the rail, in a place and in the vicinity of crest curve - profile fold (change in gradient, profile bends) and engineering structures represented by culverts and railway level crossings,

- vertical irregularities especially in linear structures, near and at the values climaxes of CWR creeping of the rails, engineering structures represented by culverts and railway level crossings,
- horizontal irregularities especially in linear structures located near and on railway level crossings and transition curves. The research showed an escalation of this parameter also in the structural assembly based on wooden sleepers - applied in the CWR protection spans.

Creeping of the rail can lead to a build-up of stress in certain sections of the railway track and consequently in deformation. Sensitivity of geometric parameters in the sustainability development of Continuous Welded Rail identifies CWR locations requiring maintenance intervention. Especially the defects in the parameters of horizontal and vertical irregularities, even more so when they are in interaction with the symptoms of creeping of the rail.

The most sensitive CWR parameters, which are especially represented by vertical and horizontal irregularities and the gradient of track gauge characterise its newralgic state. Being of considerable importance in its maintenance, exploitation and life cycle management. As a result, it provides support in the design and planning of repairs.

At the same time, it is important to be aware that the geometrical parameters studied in this publication, especially: gradient of track gauge, irregularities of track rails in the horizontal plane, irregularities of track rails in the vertical plane, are among the most sensitive parameters in the sustainable development of Continuous Welded Rails. Especially on curvilinear segment.

IV. CONCLUSIONS

In our current publication, we have distributed research in contributing to the scientific discipline of civil and transportation engineering in the subject area of sensitivity of geometric parameters in the sustainability of Continuous Welded Rail. Based on real measurement data of six geometric parameters (H: track gauge, gradient of track gauge, irregularities of track rails in the horizontal plane and

V: cant, twist, irregularities of track rails in the vertical plane) the most sensitive parameters in the sustainability of CWR were defined. The research included defects in each parameter on real objects – a specific case study, characterized by different geometric characteristics in plan and profile and operational elements of transportation engineering. The conducted research confirmed the reasonableness of the hypotheses. An explicit case study provided the determination of the segments CWR the most vulnerable to exceeding permissible deviations and selection of the most sensitive geometric parameters CWR.

The research answered that the insufficient maintenance of gravel prism layer formation in CWR is reflected in the parametric defect especially the gradient of track gauge. For the most sensitive parameters in sustainability CWR belong in terms of the plane H: gradient of track gauge and horizontal irregularities, while in the plane V: vertical irregularities. These escalate especially on curvilinear segments, requiring more significant maintenance capacity.

The conducted research also confirmed that both surveying and diagnostic knowledge, as well as measurement technologies (use of advanced measuring instruments and methods and techniques) combining scientific and practical knowledge, ensure the implementation of the research topic sensitivity of geometric parameters in the sustainability development of Continuous Welded Rail, having a reflection in:

- construction, operation and maintenance,
 - design and planning process,
 - innovation of surveying and diagnostic technologies,
 - life cycle management,
- rail transport infrastructure, in order to increase efficiency and availability.

The research also considered the effect of parametric defectiveness depending on the size of curve radius and curvilinear objects. The obtained research results are a novel source of knowledge and support in the analysis and evaluation of this type of construction in the discipline of civil engineering and transportation. Providing particularly new knowledge for enhancing the safety of rail transport infrastructure. This study includes monitoring CWR from the spring 2021 measurement period. This period is representative of a period of strong activity CWR, especially in terms of stability. In our next study, interaction with other measurement periods is predicted. Harmoniously focusing also on movement resistances of rail vehicles on Continuous Welded Rail Curves as a component of synergy in optimizing monitoring of CWR geometry shaped by operational processes.

ACKNOWLEDGEMENT

The article was prepared under the research subvention of AGH University of Science and Technology No. 16.16.150.545 in 2022.

AUTHOR CONTRIBUTIONS

K. Dybel: Conceptualization 50 %; Methodology 50 %; Software 51 %; Validation 50 %; Formal analysis 52 %; Investigation 50 %; Resources 50 %; Data curation 50 %; Writing—original draft preparation 50 %; Writing—review and editing 50 %; Visualization 52 %; Supervision 50 %; Project administration 50 % .

A. Kampczyk: Conceptualization 50 %; Methodology 50 %; Software 49 %; Validation 50 %; Formal analysis 48 %; Investigation 50 %;

Resources 50 %; Data curation 50 %; Writing—original draft preparation 50 %; Writing—review and editing 50 %; Visualization 48 %; Supervision 50 %; Project administration 50 % .

Both authors have read and agreed to the published version of the manuscript.

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

K. Dybel <http://orcid.org/0000-0003-2213-0562>

A.Kampczyk <http://orcid.org/0000-0001-9210-9668>

REFERENCES

- [1] B. Eller, M. R. Majid, S. Fischer, Laboratory Tests and FE Modeling of the Concrete Canvas, for Infrastructure Applications, *Acta Polytechnica Hungarica* 19 (3) (2022) pp. 9-20. <https://doi.org/10.12700/aph.19.3.2022.3.2>
- [2] V. Barna, A. Brautigam, B. Kocsis, D. Harangozó, S. Fischer, Investigation of the Effects of Thermit Welding on the Mechanical Properties of the Rails. *Acta Polytechnica Hungarica*, 19 (3) (2022) pp. 37-49. <https://doi.org/10.12700/APH.19.3.2022.3.4>
- [3] D. Németh, H. Horváth, M. R. Movahedi, A. Németh, S. Fischer, Investigation of the Track Gauge in Straight Sections, Considering Hungarian Railway Lines. *Acta Polytechnica Hungarica*, 19 (3) (2022) pp. 155-166. <https://doi.org/10.12700/APH.19.3.2022.3.13>
- [4] S. S. Nafis Ahmad, N. Kumar Mandal, G. Chattopadhyay, A Comparative Study of Track Buckling Parameters of Continuous Welded Rail, in: Proceedings of the International Conference on Mechanical Engineering 2009: ICME2009, Dhaka, Bangladesh, 2009, pp. 1-6, ICME09-AM-14
- [5] C. Li, S. Luo, C. Cole, M. Spiryagin, An Overview: Modern Techniques for Railway Vehicle On-board Health Monitoring Systems. *Vehicle System Dynamics*, 55 (7) (2017) pp. 1045-1070. <https://doi.org/10.1080/00423114.2017.1296963>
- [6] D. Lebel, C. Soize, C. Funfschilling, G. Perrin, High-speed Train Suspension Health Monitoring Using Computational Dynamics and Acceleration Measurements. *Vehicle System Dynamics*, 58 (6) (2019) pp. 911-932.
- [7] A. Sabato, C. Niezrecki, Feasibility of Digital Image Correlation for Railroad tie Inspection and Ballast Support Assessment. *Measurement*, 103 (2017) pp. 93-105. <https://doi.org/10.1016/j.measurement.2017.02.024>
- [8] M. Mrówczyńska, J. Sztubecki, A. Greinert, Compression of Results of Geodetic Displacement Measurements Using the PCA Method and Neural Networks. *Measurement*, 158 (2020) 107693. <https://doi.org/10.1016/j.measurement.2020.107693>
- [9] A. O. Shvets, Influence of Lateral Displacement of Bogies on the Freight Car Dynamics. *Наука та прогрес транспорту. Вісник Дніпропетровського національного університету залізничного транспорту* 6 (90) (2020) pp. 66 - 81. <https://doi.org/10.15802/stp2020/223519>
- [10] V. Atapin, A. Bondarenko, M. Sysyn, D. Grün. Monitoring and Evaluation of the Lateral Stability of CWR Track. *Journal of Failure Analysis and Prevention* 22 (2022) pp. 319-332 <https://doi.org/10.1007/s11668-021-01307-3>
- [11] A. Kampczyk, K. Dybel, The Fundamental Approach of the Digital Twin Application in Railway Turnouts with Innovative Monitoring of Weather Conditions. *Sensors* 21 (17) (2021) 5757. <https://doi.org/10.3390/s21175757>
- [12] N. Mirković, L. Brajović, Z. Popović, G. Todorović, L. Lazarević, M. Petrović, Determination of Temperature Stresses in CWR Based on Measured Rail Surface

- Temperatures. *Construction and Building Materials*, 284 (2021) 122713.
<https://doi.org/10.1016/j.conbuildmat.2021.122713>
- [13] J. Kukulski, P. Gołębiowski, J. Makowski, I. Jacyna-Gołda, J. Żak, Effective Method for Diagnosing Continuous Welded Track Condition Based on Experimental Research. *Energies*, 14 (10) (2021) 2889.
<https://doi.org/10.3390/en14102889>
- [14] T. F. Brustad, R. Dalmo, Railway Transition Curves: a Review of the State-of-the-Art and Future Research. *Infrastructures* 5 (5) (2020) 43.
<https://doi.org/10.3390/infrastructures5050043>
- [15] N. Hasan, Threshold Radius of a Ballasted CWR Curved Track: Curve Classification. *Journal of Transportation Engineering, Part A: Systems*, 143 (7) (2017) 04017026.
<https://doi.org/10.1061/JTEPBS.0000054>
- [16] G. R. Doyle Jr., M. A. Thomet, Effect of Track Geometry and Rail Vehicle Suspension on Passenger Comfort in Curves and Transitions. *Journal of Manufacturing Science and Engineering* 99 (4) (1977) pp. 841-848.
<https://doi.org/10.1115/1.3439360>
- [17] F. Pospischil, Längverschweißtes Gleis im Engen Bogen: Eine Betrachtung der Gleislagestabilität. *Studia Universitätsverlag Innsbruck*, 1. Auflage (2015).
- [18] Instruction for Measuring, Researching and Assessing Track Condition Id-14 (D-75). 2010. Available online: https://www.plk-sa.pl/files/public/user_upload/pdf/Akty_prawne_i_przepisy/Instrukcje/Podglad/Id-14.pdf (accessed on 22 April 2022).
- [19] Technical Conditions for Maintaining Track Surface on Railway Lines Id-1 (D-1). 2015. Available online: https://www.plk-sa.pl/files/public/user_upload/pdf/Akty_prawne_i_przepisy/Instrukcje/Wydruk/Warunki_tech_niczne_Id-1_ujednolic..pdf (19 April 2022).
- [20] Instruction for railway superstructure diagnosis Id-8. 2005. Available online: https://www.plk-sa.pl/files/public/user_upload/pdf/Akty_prawne_i_przepisy/Instrukcje/Wydruk/Id/Id-8_WCAG.pdf (accessed on 22 June 2022).
- [21] K. Towpik, Tor Bezstykowy – Zagrożenia, Diagnostyka, Utrzymanie, *Prace Naukowe Politechniki Warszawskiej. Transport*, 114, (2016) pp. 417-426.
- [22] J. Uren, W. F. Price, Transition Curves, in: *Surveying for Engineers*. English Language Book Society Student Editions, Palgrave, London, 1985.
https://doi.org/10.1007/978-1-349-07348-1_10
- [23] A. Basak, E. Nowak, Dynamics of Railway Vehicles Movement on Transition Curves, *Advances in Science and Technology Research Journal*, 15 (4) (2021) pp. 21-29.
<https://doi.org/10.12913/22998624/142237>
- [24] A. Kampczyk, Pomiary odległości wewnętrznych płaszczyzn kół zestawów kołowych. *TTS Technika Transportu Szybowego*, 12, (2015), pp. 31-39.
- [25] K. Dybeł, A. Kampczyk, Movement Resistances of Rail Vehicles on Continuous Welded Rail Curves. In: H. Kratochvílová, R. Kratochvíl (eds.) *Proceedings of IAC 2022 in Prague, International Academic Conference on Transport, Logistics, Tourism and Sport Science (IAC-TLTS)*. Prague, Czech Republic, Czech Institute of Academic Education, IAC202205002, 2022, pp. 78-87.



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.

The maintenance of the railway superstructure and its influence on the track geometry of regional line

Janka Šestáková^{1,*}, Alžbeta Pultznerová¹, Martin Mečár¹

¹ Department of Railway Engineering and Track Management, Faculty of Civil Engineering, University of Žilina
Univerzitná 8215/1, 010 26 Žilina, Slovak Republic

*e-mail: janka.sestakova@uniza.sk

Submitted: 19/05/2022 Accepted: 14/07/2022 Published online: 09/08/2022

Abstract: The results of the railway track quality assessment, obtained as part of the diagnostics during its operational phase, are used by the construction manager to plan repair activities. The aim of the railway infrastructure manager is to maintain the longest possible good condition of the structure, most often represented by a stable track geometry quality. In the case of low economic efficiency of quality assurance of the structure through the improvement of diagnosed parameters, according to the monitored factors in the diagnostics results it is possible to decide on the operability of the structure or its individual structural elements. The interval between repairs of determining geometrical parameters or representative quality indicators, shortened to technically, technologically, and economically inefficient time, indicates the end of life of the component or structural unit and it is necessary to plan and perform its replacement. In many cases, the structure continues to operate at the final phase of its life, for example, due to financial constraints. The infrastructure manager continues to carry out regular diagnostics and then plans and carries out routine maintenance activities to ensure a safe and reliable track. The article deals with the issue of interval diagnostics and related effects of corrections of the track geometry quality of the selected section of the regional railway line with a continuously repaired railway superstructure. Attention is paid to determining the degradation rate of track geometry quality in relation to achieving the limit values of quality indicators and the efficiency of corrective maintenance.

Keywords: diagnostics; life cycle; quality index; railway track geometry interval diagnostics; correction of the track geometry quality

I. INTRODUCTION

Individual structural units, parts or elements of the railway track are described by sets of defined evaluated parameters, characteristic for structural groups of the railway track. In the operational phase of the service life of the structure, the following parameters characterizing the quality of the railway superstructure are verified as a matter of priority: track geometry and geometric and material properties of the components of the track skeleton and the ballast bed. (**Fig. 1**).

The railway track for railway vehicles consists of two parallel rails fastened at the prescribed distances to the rail supports. If the track is not in the designed geometric position, irregularities cause dynamic effects in the dynamic system track - vehicle, i.e. vibrations both track and vehicle, which causes a decrease in running comfort and a gradual deterioration of the quality of the track structure.

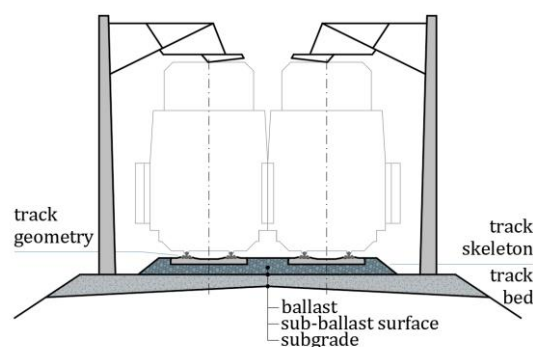


Figure 1. Structure of railway track

1. Traffic and non-traffic loading of the track

In addition to the traffic load of static and dynamic forces, the structure of the railway track is also

stressed by the load related to railway operation and repair work and the load imposed by climatic influences.

The traffic load resulting from the interaction of the rolling stock and the track (wheels of the rolling stock and the rail) cause the highest stress of the rails, or switches in railway turnouts, the load-bearing capacity of which determines the load-bearing capacity of the entire structure of the railway superstructure or turnouts. The rail is directly (centrically or eccentrically) loaded by vertical wheel forces (Q , in the z -axis direction), horizontal (guiding) forces transverse to the rail axis (Y , in the y -axis direction) and wheel forces parallel to the rail axis (T , in the x -axis direction). The rail is also loaded by the normal force N caused mainly by changes in the rail temperature. (Fig. 2).

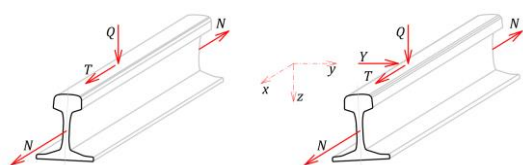


Figure 2. Forces loading the rails

The vertical load, represented by the vertical wheel force Q , is composed of quasi-static components and a dynamic component. The loading of the outer (elevated) rail in the horizontal transverse direction by the total transverse wheel force Y is also composed of quasi-static components and a dynamic component [1]. The vertical wheel force causes the bending stress of the foot and the rail head, while the stress of the rail foot is not fundamentally affected by the eccentricity of the force Q and the magnitude of the force Y [2]. The combination of the loading of the rail with the forces Q and Y causes normal stresses in the rail. Residual stresses from production, rail handling, contact stresses on the rail head and stresses from the dynamic shocks of the rail vehicles also contribute to the rail stress. The position of the wheelset of a moving rail vehicle causes an uneven distribution of the load of variable values on the rails.

Loading the track structure with the wheels of rolling stock is an iterative process. Cyclic stress causes fatigue failure of loaded components. The total dynamic track stress can be determined as the product of the static force and the dynamic coefficient, depending on the line speed and track quality [3]. In addition to the wheel forces, the forces acting on the fastening and rail support elements arising from the movement of the rolling stock, in which a vertical 'wave' is generated in front of and behind the wheel in approximately two-meter sections, also contribute to the vertical loading of the track. In addition to temperature changes (especially in continuous welded rails), changes in normal

forces in the rails also cause changes in the shape of the rail grate, such as slipping of the track or creeping of rails as a result of deficiencies in fastening nodes and ballasting of the track. The acceleration and braking of the rolling stock cause horizontal longitudinal force T . The limit values for the load on the railway are set by [4].

2. The reaction of the structure to loading

The structure of the railway track withstands the stress of traffic and non-traffic loads *due to its material properties and the interaction of structural elements (resistances) against force actions*. The condition of the structure is influenced by the current structural and geometric arrangement of the track (track geometry), the system of the railway superstructure, the structure of the railway substructure, the type and intensity of railway traffic, the condition, maintenance [5] and speed of railway vehicles, or the quality of railway superstructure and substructure repairs. After exceeding the limit value of the relevant parameter, the railway track responds to the load by deteriorating quality and the occurrence of the defect. From the point of view of actually performed activities of finding and evaluating operational quality, defects and imperfection of the spatial position of the track, track geometry and material and geometric quality of structural elements of the track grate and track bed are registered in the structure of the railway superstructure.

Horizontal transverse forces affect the occurrence of track alignment faults. In the spatial position, these forces cause the track to move in the direction of their action, while the mutual distance of the track axes on multi-track lines and the distance from fixed barriers changes. In the geometry of the track, there is a change in curvature and sudden changes in the smoothness of the track represented by the deviation of the actual versine from the design versine value. *The vertical forces cause elastic or permanent deformations of the track level*, manifested by a lowering of the vertical track alignment of one or both levels of the rails. Deformations can form in continuous or short isolated sections. In the spatial position, the changes manifest themselves as a continuous (most often uneven) decrease, while the position of the cross-section and the mutual position of the track and the traction line change. In the geometry of the track, there are decreases of rail joints and low point of the track, changes in the superelevation – cannot of the track, or also twist of the track.

Dynamic loading of the track skeleton and fitting defects, or maintenance of the track skeleton are the cause of changes in the track gauge. The most frequent reason for the reduction of the track gauge is the rolling of the rail heads, deflection of the sleepers, and uneven slipping of the rails. Increasing

the track gauge causes wear and loosening of the fasteners, lateral wear of the rails, uneven pushing of the baseplates into the rail supports, deflection of the sleepers, incorrectly designed track cannot, poor quality of the track skeleton assembly and the condition of the railway vehicles wheels.

II. RELATED WORKS

The quality of a railway track is represented by all its prescribed technical and ecological characteristics during the entire lifetime: in the construction phase and the operation phase. In operation, the quality of the railway track is described by the state of the parameters of the track geometry, possibly also by the state of the geometry and material of structural elements or the whole structure. The behavior of the structure, which is assessed based on its quality, is represented by a value of the quality index for the assessed section. The determination of the quality index of the track section is affected by the structure of the data that enters the evaluation process. The structure and content of the data affect the accuracy of the prediction of the future development of the construction condition. The basic group of data is information about the geometry of the track, the geometry of the rail profiles, the mechanical characteristics of the substructure and the characteristics of the rolling stock and their response to the condition of the railway track. The quality of the track geometry is defined by the values of deviations from the reference geometric characteristics of the prescribed parameters in the alignment and level of the track, and thus it is determined from the diagnostics data of railway track spatial position (it is clearly defined in the plan by co-ordinates) and track geometry (geometrical position arrangement: track gauge, the relative level of rails: cant, cant gradient, the relative gradient of rails, level of the track and alignment and profile of line). The quality of the track geometry is significantly influenced by the condition of the rails with their material and geometric parameters. The rail head profile has a direct impact on the track geometry if unevenness and shape changes are formed in the running surface and top surface of the railhead. These rail defects are caused by traffic load and have the character of a deviation from the reference shape with various depths and lengths.

The quality of a railway track changes in the individual phases of its life cycle. The initial phase of the life cycle is represented by the parameters achieved by construction activities. In the operational phase of the railway, its condition is affected by the effects of traffic and non-traffic loads, which are manifested by permanent changes in the geometric parameters and material characteristics of the structure and structural elements.

Throughout the lifetime of the railway track, parameters representing the state of the structure are recorded, evaluated, and analysed (diagnostics). In relation to the standardized values of the prescribed parameters, the quality of the railway track repeatedly goes through phases of degradation and rehabilitation. Evaluation and analysis of data representing the degradation of railway track quality influence decisions on optimal intervals for determining the current state of the structure [6], estimates of residual life, determination of life cycle costs of the whole structure [7], [8], or its components [9], [10], [11] and to predict the appropriate time and format of quality rehabilitation. While defects occur in the initial phase of the life cycle (consequences of structural creep, insufficient load resistance, non-compliance with production or construction technology) with a decreasing risk, the risk of defects is approximately constant in the middle of the life. At the end of its service life, the risk of defects increases and is affected by the operational condition of the structure and its parts: age of the structure, wear due to load, and insufficient or incorrect rehabilitation activities [12]. When the interval between two quality rehabilitations is reduced to technically and economically inefficient time, the service life of the structure ends, and the structure must be replaced [13] (Fig. 3).

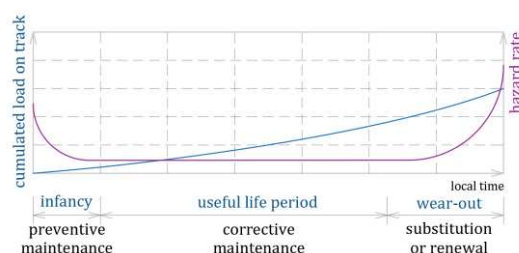


Figure 3. Cumulated load and hazard rate during the life cycle of the railway track [12], [13]

The life cycle of structures or maintenance and repair activities on a typical intensively operated main railway line loaded with average intensity operations are shown in Table 1. According to [14], the standard operating life cycle of the highest category railway line (line speed, load carried) is up to 8 years. The cleaning of the track bed is performed only in connection with other major repairs (in the track with wooden sleepers approximately every 20 years, with concrete sleepers approximately every 30 years). The service life of the rails (about 10 years) is usually half of the service life of the rail skeleton: the limiting factor is usually not the complete wear of the rail head, but the occurrence of material defects on the running surface. The replacement of the track bed is usually carried out in connection with a complete replacement of the track grate (at intervals of 20 to 30 years). The service life of turnouts is about 20 years when using wooden

sleepers and about 30 years when using concrete sleepers. In cases of extremely heavy load, it may be necessary to replace the frogs of the turnouts up to three times a year, but under average conditions, the normal life of the frog is five years.

Table 1. Typical service lives on a typical intensively operated main line [14]

Typical life cycles	Operational load	Years
tamping	40 – 70	4 – 5
grinding	20 – 30	1 – 3
ballast cleaning	150 – 300	12 – 15
rail renewal	300 – 1 000	10 – 15
timber sleeper renewal	250 – 600	20 – 30
concrete sleeper renewal	350 – 700	30 – 40
fastenings	100 – 500	10 – 30
ballast renewal	200 – 500	20 – 30
formation renewal	> 500	> 40

Railway tracks in the Slovak Republic are classified into 6 categories by the annual operating load: tracks with a load of more than 47,450 million tons per year are classified as category No. 1 and tracks with a load of fewer than 1,825 million tons per year are classified to the category No. 6 [15].

Given the limited resources of maintenance financing and the effort to reduce the time spent on diagnostics and track maintenance, it is necessary to use an efficient and effective system of maintenance linked to diagnostics [16], [17]. Targeting maintenance to meet safety limits and achieve cost-effectiveness directs maintenance strategies from corrective to preventive maintenance, which places extremely high demands on the quality of diagnostic outputs, especially in the field of evaluation, analysis, and prediction [18], [19], [20], [21].

The railway infrastructure manager can increase the cost and time efficiency of maintenance by using maintenance planning scenarios that are based on analyses of structural failure risk, i.e., occurrence of a critical error [22]. A suitable model of the degradation of the parameters of a structure or structural elements makes it possible to determine the limit value of the operating time or accumulated traffic load in the event of a structural failure. It is possible to use analytical tools of safety limits and optimal time of maintenance in the management and maintenance of railway structures. To prevent the risks of failure, modern concepts of reliability, availability, maintainability, and safety (RAMS) analysis are used for railway tracks. Analyses help to minimize the risks of failure to an acceptable level [23].

One of the ways to ensure the safety of the railway track and the comfort of running railway vehicles is to maintain the high quality of its geometry. The low quality of the track geometry can directly or indirectly result in safety problems, speed reduction, traffic reduction or interruption, higher maintenance costs and a higher degree of quality degradation of affected structures (rails, turnouts, crossings, etc.) and railway vehicles. If the correct track geometry maintenance strategy is not selected, the quality of the structure may deteriorate above a defined level (e. g. intervention limit – *IL*), leading to a higher frequency of track geometry repair (tamping) and consequently higher maintenance costs [24], [25].

In order to make an effective decision on how to repair the railway line, it is necessary to have representative data from the entire evaluated section. Signals of track geometry parameters are at the first level (diagnosed quality indicators) of diagnostics data. Their structure is shown, for example, in Chapter III, Part 1 of this article. The quality of the geometric parameters of the track affects the quality of each evaluated section, which is expressed by the diagnostics data at the second level (calculation, valuation): track quality index (*TQI*) [26]. *TQI* is a combination of diagnosed quality indicators, and several approaches can be used to calculate it, which were developed for different railway infrastructure administrations [27] and use numerical calculation methods. The authors have been dealing with the issue of verifying the quality of railway track geometry since 2012. The main goal of monitoring the parameters of track geometry is to verify the rate of degradation of its quality and to model future quality development. The experiments are focused on the transition areas between the track bed structure and the slab track on the main lines of the Railways of the Slovak Republic (ŽSR) [28], [29] and on monitoring the quality development of selected sections of railway lines of regional importance. In this article, the authors focus on the evaluated diagnostics of the track geometry, the structure of which is in relation to the technical quality and technological efficiency of maintenance and repairs in the final phase of the life cycle.

III. METHODS

Configuration of the tested section of railway line no. 162 in the ŽSR Lučenec – Utekáč network is shown in Fig. 4. For software evaluation, the beginning of the tested section is simply identified: km 0.000 000 = rkm 24.550 000.

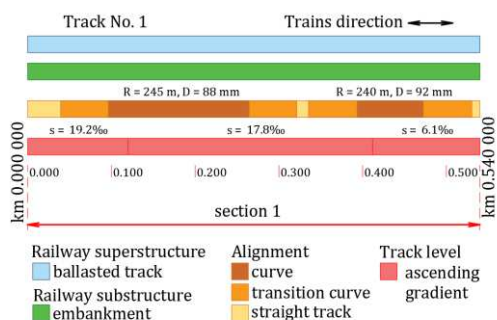


Figure 4. Segmentation of the tested section

The alignment of the section consists of two curves with transition curves and adjacent straight sections. The grade of track level of the whole section rises in the direction of the stationing. The values of the radii of the curves (R), their cants (D) and gradients of the track level (s) are in Fig. 4.

The tested section was put into operation in 1958. It is currently included in the SR1 speed range and trains run on it at a maximum speed (V) of $60 \text{ km}\cdot\text{h}^{-1}$ with some sections with permanently reduced speed due to lower quality of track geometry and level crossings with road communications without safety and signalling. The line is used mainly for passenger train transport, namely light diesel class series 813 and 913. That investigated section is a part of the railway line, which is included in category No. 6 with an annual operating load of fewer than 1.825 million tons.

The structural composition and condition of the track skeleton in the track bed corresponds to the age of the structure: jointed track with A/75 rails 20 m long, concrete sleepers (PAB or SB2 with division 'c') in a combination of single or short sections with wooden sleepers, The fastening of the rails is stiff with clip baseplates. The structural elements of the railway superstructure show defects and imperfections (Fig. 5 to Fig. 8) typical for the structure on which corrective maintenance is applied [30].



Figure 5. Railway superstructure of the tested section



Figure 6. Defect of a wooden sleeper



Figure 7. Pushed baseplate



Figure 8. Lip on the rail head

1. Data collection

The diagnostic of the monitored section focuses on the collection of track geometry data, the calculation of prescribed quality indicators and the evaluation of the quality of track geometry parameters. As part of the experimental verification, five measurements ($M1$ to $M5$) have been carried out at approximately half-yearly intervals since 2019.

Measurements are performed with an electronic hand measuring trolley KRABTM-Light, which captures the track condition without load. Data collection is performed by a continuous method – reading and recording of measured quantities every 0.25 m, while the track structure is not under traffic load during the measurement. The measurement method is in the conditions of the ŽSR the main method allowed for tracks operated at a maximum speed of $120 \text{ km}\cdot\text{h}^{-1}$. During the measurement, all required values of the track geometry are read and recorded at the same time, which defines the method

as a complex method. The sensors of the measuring device are in contact with the structural elements (rails) of the measured structure during the duration of sensing the quantities.

The track geometry in terms of [26] is represented by:

- quantities of the geometric arrangement of the track:
 - deviations from the designed position of the track alignment (*Al*) (unit: mm),
 - deviations from the designed level of the rail position (*Tp*) (mm),
- track design values:
 - track gauge (*Ga*) (mm),
 - track cant (*Ct*) (mm)

The quantities *Al* and *Tp* characterize the running track of rolling stock, which is represented by spatial curves, which are also in the whole range of wavelengths. It is only possible to measure them indirectly - by determining the smoothness of the curves that the alignment and vertical profile of the track create. The evaluation of the fluidity of the curves is performed on the principle of measuring the versines on the chords. The quantities *Ga* and *Ct* are direct measures in the respective cross section of the track, they are directly (absolutely) measurable as the so-called real geometry, in the whole range of wavelengths $\lambda \in \langle 1\text{m}, \infty \rangle$. From the track gauge, the track gauge change is calculated per 1 m of the track (*Ga/m*; mm/m). From the cant values, it is possible to derive the twist of the track (*Tw*) on the bases of the prescribed lengths (for ŽSR 3.00, 6.00 or 12.00 m (mm/m)) [31].

2. Evaluation of track geometry quality

In the conditions of the ŽSR, the operating condition of the railway track is primarily evaluated and presented by the condition (quality) of the track geometry, the condition of individual structural elements of the track skeleton, or the ballast bed.

In [31], limit levels of permissible deviations from nominal or projected values of geometric quantities or limit values of quantities (highest values of local defects parameters) are defined for track geometry quality evaluations.

The Alert Limit (*AL*), Intervention Limit (*IL*) and Immediate Action Limit (*IAL*) are intended for the evaluation of the operational quality of the railway track, while the achievement or exceeding of *IL* levels or *IAL* is being monitored. The values depend on the speed range in which the diagnosed structure is categorised. The maintenance time also depends on the inspection interval and the speed of the evolution of the defect:

- operating deviations from the designed or prescribed value of a geometric quantity during the operation of a railway line:

- Alert Limit: if the set value is exceeded, the state of the track geometry variables must be assessed and taken into account in the planning of repair and maintenance activities,
- Intervention Limit: if the set value is exceeded, maintenance work must be carried out so that the operating deviation is not exceeded before future inspections,
- limit operating deviations from the designed or prescribed value of the track geometry quantity during track operation, which should not be exceeded and are defined as the Immediate Action Limit; if this deviation is exceeded, structure and transport measures must be taken to reduce the risk to an acceptable level.

Limits for *SR1* ($V \leq 60 \text{ km}\cdot\text{h}^{-1}$) during operation are in **Table 2**.

Table 2. Limit deviations and values of track geometry parameters during operation for *SR1* [32]

Parameter	Value					
	AL		IL		IAL	
<i>Ga</i> (mm)	-5	30	-7	32	-9	35
<i>Ga/m</i> (mm/m)	not evaluated		5		6	
<i>Ct</i> (mm)	-13	13	-15	15	-20	20
<i>Tp</i> (mm)	-15	15	-20	20	-28	28
<i>Al</i> (mm)	-15	15	-18	18	-20	20
<i>Tw</i> ₁ (mm/3.0 m)	12		15		18	
<i>Tw</i> ₂ (mm/6.0 m)	16		20		24	
<i>Tw</i> ₃ (mm/12.0 m)	31		34		36	

By [33], the track geometry quality is described by the standard deviation (*SD*) of the parameter over a defined length (minimum 200 m, usually 1,000 m), the quality mark of the geometric quantity (*QM*) and the quality number of the evaluated section (*QN*) (equations (1) to (3) [32]).

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n x_i^2} \quad (1)$$

$$QN = \sqrt{1.6 \cdot SD_{Al}^2 + 0.6 \cdot SD_{Ga}^2 + 1.6 \cdot SD_{Ct}^2 + 1.6 \cdot SD_{Tp}^2} \quad (2)$$

$$QM = \frac{\ln \frac{SD}{b}}{m} \quad (3)$$

where *n* is the number of points measured each 0.25 m,

i is the designation of the measuring point, *x_i* is the dynamic component of the relevant quantity (deviation from the median in the range *D1*),

b and *m* are numeric constants determined based on *SD* statistics of the relevant

parameter and speed range.

The Railways of the Slovak Republic quality indices for SR1 during operation are in **Table 3**.

Table 3. Limit values of track quality indices during operation [33]

<i>SD of decisive quantities</i>	
SD_{AI}	3.10
SD_{Ga}	2.70
SD_{Ct}	2.90
SD_{Tp}	3.40
<i>QN of the evaluated section</i>	
QN	7.20
<i>evaluation of the track geometry condition by QM</i>	
$0 < QM \leq 2$	satisfactory
$2 < QM \leq 3$	it is recommended to design a track geometry correction to the maintenance plan
$3 < QM < 4$	it is recommended to correct the track geometry by the next inspection
$4 \leq QM \leq 6$	it is recommended to take immediate measures to arrange the safety of operation

IV. RESULTS AND DISCUSSION

Monitoring of the quality of the track geometry is put at the end of the service life of the tested section – in the near future, reconstruction is planned with a comprehensive replacement of the railway superstructure.

The evaluation of the track geometry quality takes into account in the quality indicators (SD , QM , QN) the influence of the four determinants, which are listed in Chapter II, Part 1 (AI , Tp , Ga , Ct). Since [26] sets the alignment (AI) and vertical level of rail (Tp) values of the track (rails) as determining variables for maintaining the reliability and safety of railway operation, some evaluation outputs are processed in this article only for these variables (**Table 4**, **Fig. 9** to **Fig. 12**).

The period of monitoring carried out so far is limited by the events of No. 1 to No. 7 (**Table 4**) and contains 2 activities related to the rehabilitation of the quality of the track geometry:

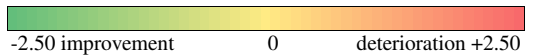
- activity No. 1: maintenance focused on manual elimination of track geometry defects detected by the recording car of the ŽSR, which captures the track condition under load (section quality assessment day 0, rehabilitation phase No. 1),
- activity No. 6: tamping of the track with an automatic lifting and levelling tamping machine, backfilling of material of the track bed and its adjustment to the prescribed profile and a unique replacement of the track structure components (day 774 of the section quality assessment, rehabilitation phase No. 2).

The determination of the current state of the track by diagnostics of its geometry in the cases of

measurements $M1$ and $M5$ took place shortly after the previous maintenance by the measurement trolley, which captures the track condition without load: 66 or 14 days and the interval between individual measurements is 160 to 197 days.

Table 4. Improvement or deterioration of track quality indexes between measurements

Activity	Days from action No. 1	Improvement (-) deterioration (+)		
		SD_{AI}	SD_{Tp}	QN
No. 1 Manual maintenance	0			
No. 2 $M1$	66	-0.13	-0.20	-0.38
No. 3 $M2$	254	+0.28	+0.14	+0.35
No. 4 $M3$	414	-0.12	+0.03	-0.11
No. 5 $M4$	611			
No. 6 Machine maintenance	774	-1.14	-1.31	-2.12
No. 7 $M5$	788			



It follows from the above that the $M1$ measurement is placed almost at the beginning of the 1st degradation phase of the track geometry quality, which ends with activity No. 6. The $M5$ measurement is placed at the beginning of the 2nd degradation phase, which currently lasts. Development of improvement or deterioration of the quality indicator between individual measurements in **Table 4** presents a constant level of quality, represented by indicators SD_{AI} and SD_{Tp} .

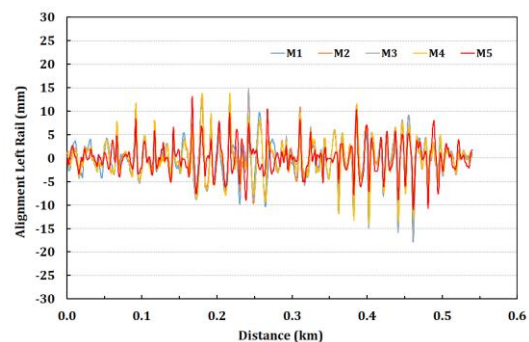


Figure 9. Deviation of alignment of left rail

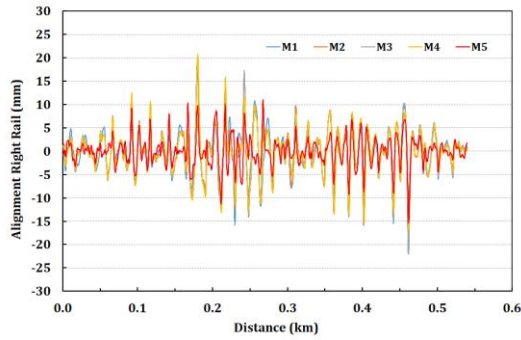


Figure 10. Deviation of alignment of right rail

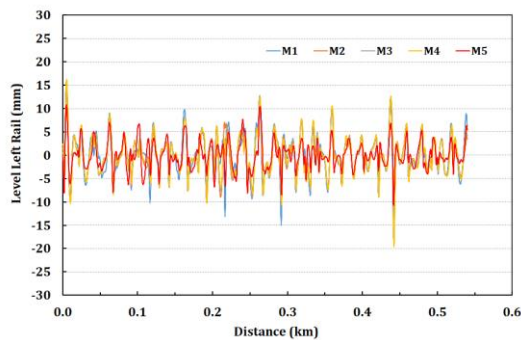


Figure 11. Deviation of level of left rail

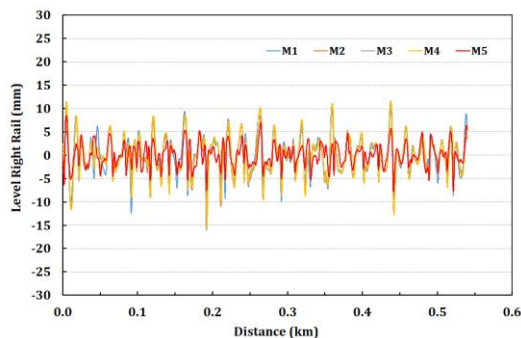


Figure 12. Deviation of level of right rail

The values of their improvement / - / or deterioration / + / between measurements are from -0.20 to +0.58 mm) and QN (-0.38 to +0.35). After the corrective activity at the level of the machine tamping and the related improvements of the track bed, a significant improvement in the quality of the track geometry is detected. The track rehabilitation phase is therefore quantified by a reduction (improvement) in SD_{AI} of 1.14 mm, SD_{Tp} of 1.31 mm and a QN was improved by 2.12.

As can be seen from the graphs in Fig. 9 to Fig. 12, the deviations of the determining quantities from the projected position AI and Tp exceed the values of the permitted deviations AL , IL or IAL only in short solitary sections:

- in the 1st degradation phase, a maximum of 5 sections (AI), 2 (Tp) in the level of AL , or 1 (AI) at the IAL level,

- in the 2nd degradation phase, 1 section (AI) in the AL level.

Clear display of the total length of sections with exceeded permissible deviations of all measured values of the track geometry at AL levels, IL and IAL , and the ratio of the length of the sections with defect to the total length of the test section is shown in Fig. 13. The figure visualizes the course of the track geometry quality in terms of the length of defects in the 1st degradation phase between the 0th and 1st rehabilitation phase (vertical dashed lines) and the initial track geometry quality in the 2nd degradation phase of the structure quality.

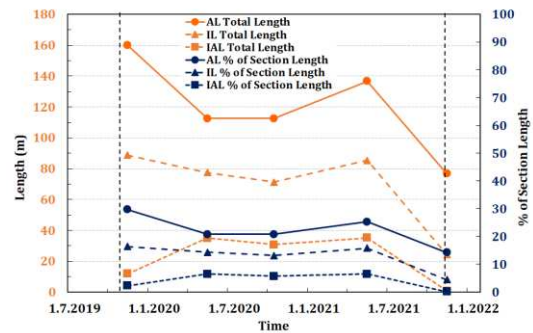


Figure 13. Total length of track geometry defects and percentage of examination section length

The length proportion of the individual rail geometry values in the total length of the section with defects is shown separately for the level of AL (Fig. 14), IL (Fig. 15) and IAL (Fig. 16). In this view, the worst variable is Ct , which is the best variable in terms of the variance of the mean value (Fig. 17).

With the support of the outputs of the measuring device, we can assume that within the repeated cycles of track geometry correction, the cant was gradually built by tamping in the long section of the track, exceeding the projected cant above the value of permissible deviations. Defects of parameters AI and Ct were corrected by tamping after $M4$. The Ga/m defects were corrected by individual replacing of sleepers after $M5$.

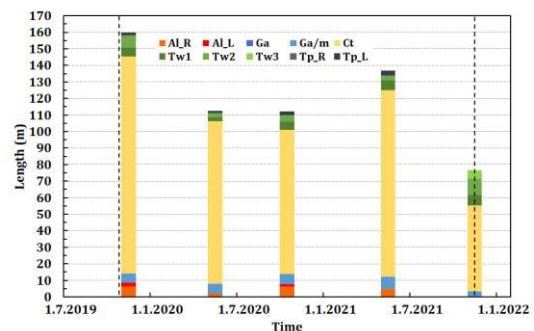


Figure 14. Total length of track geometry defects (alert limit)

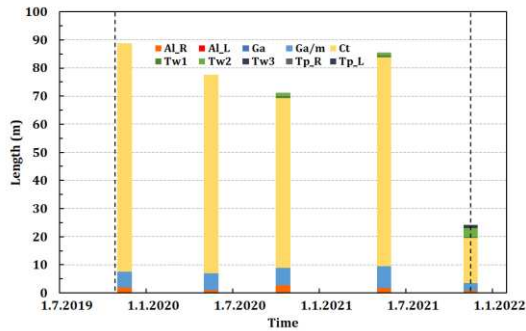


Figure 15. Total length of track geometry defects (intervention limit)

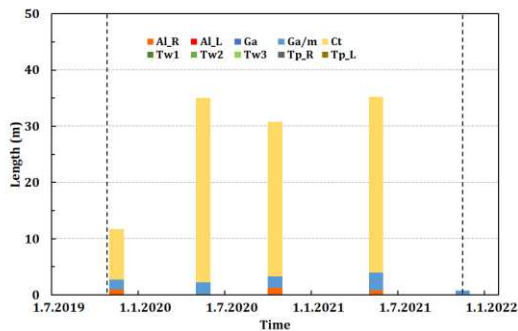


Figure 16. Total length of track geometry defects (immediate action limit)

The frequency of occurrence of defects determining the quantities Al and Tp does not indicate a significant deterioration in the quality of the alignment and vertical profile of the track position of the tested section. The variance of the values of the deviations of these quantities from the average value by means of SD_{Al} , SD_{Tp} (Fig. 17) and QM_{Al} (Fig. 18) already places the section in a category with unsatisfactory quality. Such a combination of quality indicators indicates the multiple occurrences of deviations in the direction and elevation of the rails from the design position – for example at rail junctions – without exceeding the permissible deviation.

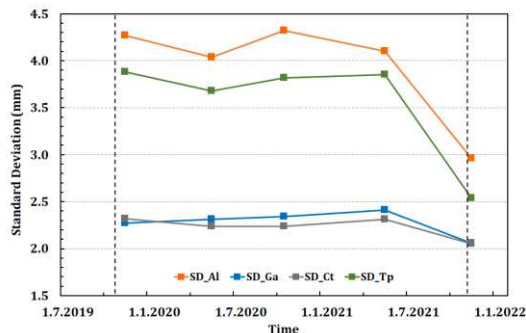


Figure 17. Standard deviations of track geometry parameters

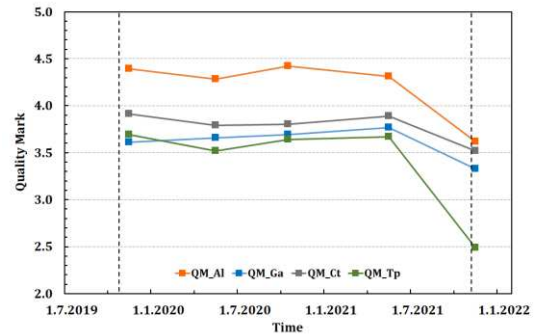


Figure 18. Quality marks of track geometry parameters

Based on the determined values of the quality number, it is possible to describe the development of the quality of the track geometry of the tested section (Fig. 19) as follows:

- the railway structure maintains the trend of not deteriorating the quality of the track geometry at the end of its service life under favourable operating conditions (intensity and mode of transport), but
- manual maintenance, implemented in selected short sub-sections, is not sufficient in the railway structure for the long-term elimination of track geometry defects and improvement of track geometry quality indicators above-defined limits, especially in determining parameters of track geometry Al , Tp ,
- machine maintenance at the level of the rail tamping and supplementing of the rail bed material and the unique replacement of the track skeleton components rehabilitates the rail geometry to the prescribed quality level, but the results of the following monitoring activities in the 2nd degradation phase will provide information on the durability of such improvement.

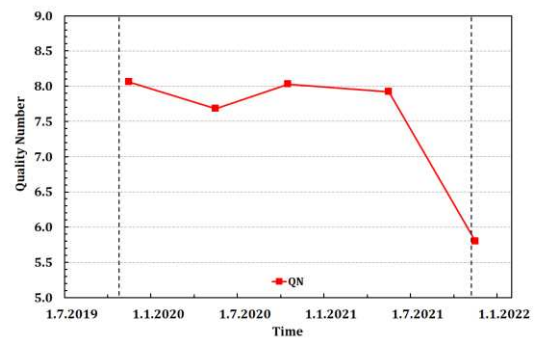


Figure 19. Track quality indexes

V. CONCLUSIONS

Monitoring of the track geometry, which forms the track structure at the end of the service life phase, focuses on verifying the effectiveness of maintenance that has the character of corrective

maintenance – i.e. is a response to defects detected by the measurement trolley. In the monitoring period (since September 2019), 5 measurements of track geometry values and 2 phases of quality rehabilitation – manual and machine maintenance (tamping) were performed, accompanied by a supplementary and unique replacement of track skeleton components. The results of the measurements show that the quality of the track geometry is maintained at approximately the same quality at the end of its service life under favourable operating conditions (intensity and type of transport) during the entire 1st phase of degradation. Parameters and efficiency of manual maintenance (0th rehabilitation phase), implemented in selected short sections, did not improve the quality of track geometry below the prescribed limit values of parameters and quality indices and this method of maintenance was not sufficient to eliminate track geometry defects and improve track geometry quality indicators, especially in directional and elevation parameters of the track and rails. Quality assessment of machine maintenance (1st rehabilitation phase) showed a significant improvement in the quality of the track geometry at the beginning of the 2nd degradation phase, up to the prescribed quality level.

Verification of the durability of the improvement in the track geometry quality will be the next research of the investigated section. It is assumed to create and evaluate a prediction model of quality degradation to determine the optimal interval of track geometry maintenance.

ACKNOWLEDGEMENT

The publishing of this paper was supported by the grant VEGA 1/0643/21 Analysis of Spatial

REFERENCES

- [1] C. Esveld, Modern Railway Track, Digital Edition, Version 3.8, MRT-Productions, Zaltbommel, 2016, ISBN 978-1-326-05172-3.
- [2] O. Plášek, P. Zvěřina et al. Railway Constructions. Railway Substructure and Superstructure, 1st Edition, Akademické nakladatelství CERM, s. r. o., Brno, 2004, ISBN 80-214-2621-7, in Czech.
- [3] B. J. Van Dyk, J. R. Edwards et al. Evaluation of Dynamic and Impact Wheel Load Factors and their Application in Design Processes, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 231 (1) (2017) pp. 33-43. <https://doi.org/10.1177/0954409715619454>
- [4] Testing and Approval of Railway Vehicles from the Point of View of their Dynamic Behaviour – Safety – Track Fatigue – Ride Quality, 3rd edition, UIC Code 518:2005 (2005).
- [5] P. Zvolensky, J. Grecnik, J. Gallikova, ECM Certification Changes and Procedures under the new EU Commission Regulation 2019/779, 24th International Conference on Current Problems in Rail Vehicles, Zilina, Slovakia, Sep 17-19, 2019, EDIS-publishing Inst Univ Zilina, pp.349-356
- [6] A. Kasraei, J. A. Zakeri, Effective Time Interval for Railway Track Geometry Inspection, Archives of Transport 53 (1) (2020) pp. 53–63. <https://doi.org/10.5604/01.3001.0014.1744>
- [7] S. Bressi, J. Santos, M. Losa, Optimization of Maintenance Strategies for Railway Track-bed Considering Probabilistic Degradation Models and Different Reliability Levels, Reliability Engineering & System Safety 207 (2021),

Deformations of the Railway Track Focused on Terrestrial Laser Scanning.

This work was supported by the Grant No. 22120015. The project is co-financed by the Governments of Czechia, Hungary, Poland and Slovakia through Visegrad Grants from International Visegrad Fund. The mission of the fund is to advance ideas for sustainable regional cooperation in Central Europe.



AUTHOR CONTRIBUTIONS

J. Šestáková: Conceptualization, Experiments, Theoretical analysis, Writing, Review and editing.

A. Pultznerová: Conceptualization, Review and editing.

M. Mečár: Conceptualization, Experiments.

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

J. Šestáková <https://orcid.org/0000-0003-1983-063X>

A. Pultznerová <https://orcid.org/0000-0002-8097-9946>

M. Mečár <https://orcid.org/0000-0003-1367-0380>

- 107359, ISSN 0951-8320.
<https://doi.org/10.1016/j.ress.2020.107359>
- [8] V. Jover, L. Gaspar, Sz. Fischer, Investigation of Geometrical Deterioration of Tramway Tracks, *Science and Transport Progress*, 2 (86) (2020) pp.46–59.
<https://doi.org/10.15802/stp2020/204152>
- [9] D. Ovchinnikov, A. Bondarenko et al. Extending Service Life of Rails in the Case of a Rail Head Defect, *GRAĐEVINAR*, 73 (2) (2021) pp. 119-125.
<https://doi.org/10.14256/JCE.2860.2019>
- [10] B. Eller, Sz. Fischer, Tutorial on the emergence of local substructure failures in the railway track structure and their renewal with existing and new methodologies, *Acta Technica Jaurinensis*, 14 (1) (2021) pp. 80-103.
<https://doi.org/10.14513/actatechjaur.00565>
- [11] L. Bai, R. Liu et al. Estimating Railway Rail Service Life: A Rail-grid-based Approach, *Transportation Research Part A: Policy and Practice*, 105 (2017) pp. 54-65, ISSN 0965-8564. <https://doi.org/10.1016/j.tra.2017.08.007>
- [12] N. Lyngby, P. Hokstad, J. Vatn, RAMS Management of Railway Tracks, in: K. B. Misra (Ed.) *Handbook of Performability Engineering*, 1st Edition, Springer, London, 2008, pp. 1123–1145.
https://doi.org/10.1007/978-1-84800-131-2_68
- [13] K. Tzanakakis, *The Railway Track and Its Long Term Behaviour. A Handbook for a Railway Track of High Quality*, 1st Edition, Springer-Verlag, Heidelberg, 2013, ISSN 2194-8119. <https://doi.org/10.1007/978-3-642-36051-0>
- [14] B. Lichtberger, *The Track System and Its Maintenance*, RTR Special. Maintenance & Renewal. Special Edition, July 2007, pp. 14-22. ISBN 978-3-7771-0367-9.
- [15] *Railway Superstructure, Regulation of the Railways of the Slovak Republic TS3:2012* (2012), in Slovak.
- [16] R. Schenkendorf, B. Dutschk et al. Improved Railway Track Irregularities Classification by a Model Inversion Approach, in: *PHM Society European Conference*, 3(1).
<https://www.papers.phmsociety.org/index.php/phme/article/view/1592>
- [17] J. Kanis, V. Zitrický et al. Innovative Diagnostics of the Railway Track Superstructure, *Transportation Research Procedia. International Scientific Conference "Horizons of Railway Transport 2020"*, 53 (2021) pp. 138-145, ISSN 2352-1465.
<https://doi.org/10.1016/j.trpro.2021.02.017>
- [18] J. Neuhold, M. Landgraf, From Data-based Condition Analysis to Sophisticated Asset Management for Railway Tracks. In *Proceedings of World Congress of Railway Research*, Tokyo, Japan, 2019.
<https://www.sparkrail.org/Lists/Records/DispForm.aspx?ID=26621>
- [19] C. Vale, M. L. Simões, Prediction of Railway Track Condition for Preventive Maintenance by Using a Data-Driven Approach, *Infrastructures* 7 (3):34 (2022).
<https://doi.org/10.3390/infrastructures7030034>
- [20] X. Sun, F. Yang et al. On-Board Detection of Longitudinal Track Irregularity Via Axle Box Acceleration in HSR, *IEEE Access* 9 (2021) pp. 14025-14037.
<https://doi.org/10.1109/ACCESS.2021.3052099>
- [21] M. Sysyn, et al. Turnout Monitoring with Vehicle Based Inertial Measurements of Operational Trains: A Machine Learning Approach. *Communications - Scientific letters of the University of Zilina*, 21(1) (2019) pp. 42-48.
<https://doi.org/10.26552/com.C.2019.1.42-48>
- [22] A. Jamshidi, S. F. Roohi, Probabilistic Defect-Based Risk Assessment Approach for Rail Failures in Railway Infrastructure, *IFAC-PapersOnLine* 49-3 (2016) 073–077
<https://doi.org/10.1016/j.ifacol.2016.07.013>
- [23] S. Hidirov, H. Guler, Reliability, Availability and Maintainability Analyses for Railway Infrastructure Management, *Structure and Infrastructure Engineering*, Volume 15 (9) (2019)
<https://doi.org/10.1080/15732479.2019.1615964>
- [24] I. A. Khouy, *Cost-Effective Maintenance of Railway Track Geometry. A Shift from Safety Limits to Maintenance Limits*, Doctoral Thesis, Luleå University of Technology (2013), ISSN: 1402-1544
<https://www.diva-portal.org/smash/get/diva2:999211/FULLTEXT01.pdf>
- [25] U. Espling, *Maintenance Strategy for a Railway Infrastructure in a Regulated Environment*, Doctoral Thesis, Luleå University of Technology (2007), ISSN 1402-1544.
<https://tu.diva-portal.org/smash/get/diva2:991206/FULLTEXT01.pdf>
- [26] *Railway applications. Track. Track Geometry Quality. Part 6: Characterisation of Track Geometry Quality*, STN EN 13848-6 (73 6315):2014 (2014).
- [27] D. Setiawan M., S. A. P. Rosyidi, Track Quality Index as Track Quality Assessment Indicator. *The 19th International Symposium of FSTPT. Islamic University of Indonesia*, Oct. 11-13 2016. pp. 197-207, ISBN: 979-95721-2-19.
<https://ojs.fstpt.info/index.php/ProsFSTPT/article/download/209/205>
- [28] J. Šestáková, L. Ižvolt, M. Mečár, Degradation-Prediction Models of the Railway Track

- Quality, Civil and Environmental Engineering, 15 (2) (2019) pp.115-124.
<https://doi.org/10.2478/cee-2019-0015>
- [29] J. Šestáková, A. Matejov, A. Pultznerová, Rehabilitation of Railway Track Quality in Relation to Diagnostic Data, in: Akimov, P., Vatin, N. (Eds), XXX. Russian-Polish-Slovak Seminar Theoretical Foundation of Civil Engineering: RSP 2021. Lecture Notes in Civil Engineering, 189 (2021) pp. 197–206. Springer, Cham.
https://doi.org/10.1007/978-3-030-86001-1_23
- [30] W. Alduk, S. Marenjak, Track Maintenance at the End of Life Cycle, in: S. Lakušić (Ed.), 4th International Conference on Road and Rail Infrastructure: CETRA 2016, University of Zagreb, Zagreb, 2016, pp. 687–693.
<http://master.grad.hr/cetra/ocs/index.php/cetra>
- [4/cetra2016/paper/view/545](https://doi.org/10.2478/cee-2019-0015)
- [31] Measurement of Track Geometry. Part I. Theoretical Foundations. KRAB™ – Light Measuring Trolley. (electronic). Komerční železniční výzkum s. r. o., Prague. (2002), in Czech.
- [32] Measurement and Evaluation of the Track Gauges 1 435 mm and 1 520 mm by Measuring Trolley for Railway Track Geometry KRAB, Regulation of the Railways of the Slovak Republic SR 103-7 (TS):2017 (2017), in Slovak.
- [33] Railway Applications. Track. Part 2: Acceptance of Construction Works, Maintenance Works, and Assessment of Service Condition Track Gauges 1 435 mm, STN 73 6360-2:2015 (2015), in Slovak.



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.

Simulation and Genetic Algorithms to Improve the Performance of an Automated Manufacturing Line

Patrick Ruane^{1,2,*}, Patrick Walsh², John Cosgrove²

¹ Johnson & Johnson Vision Care
Rivers, V94 N732 Limerick, Ireland

² Technological University of the Shannon
Moylish, V94 EC5T, Limerick, Ireland

*e-mail: patrick.ruane@tus.ie

Submitted: 27/06/2022 Accepted: 29/08/2022 Published online: 31/08/2022

Abstract: Simulation in manufacturing is often applied in situations where conducting experiments on a real system is very difficult often because of cost or the time to carry out the experiment is too long. Optimization is the organized search for such designs and operating modes to find the best available solution from a set of feasible solutions. It determines the set of actions or elements that must be implemented to achieve an optimized manufacturing line. As a result of being able to concurrently simulate and optimize equipment processes, the understanding of how the actual production system will perform under varying conditions is achieved. The author has adopted an open-source simulation tool (JaamSim) to develop a digital model of an automated tray loader manufacturing system in the Johnson & Johnson Vision Care (JJVC) manufacturing facility. This paper demonstrates how a digital model developed using JaamSim was integrated with an author developed genetic algorithm optimization system and how both tools can be used for the optimization and development of an automated manufacturing line in the medical devices industry.

Keywords: Digital Model; Digitalization; Genetic Algorithm; JaamSim; Optimization; Simulation

I. INTRODUCTION

Digitalization in manufacturing is the conversion of information into digital format, the integration of this digital data and technologies into the manufacturing process and the use of those technologies (eg: simulation, optimization) to change a business model to provide new revenue and value-producing opportunities. Digitalization may be seen as the increased generation, analysis, and use of data to improve the efficiency of the overall manufacturing system. Digital manufacturing technologies, such as simulation models, have been considered an essential part of the continuous effort towards improving the performance of automated manufacturing equipment and processes. Optimization seeks the maximum or minimum value of an objective function corresponding to variables defined in a feasible range or space. More generally, optimization is the search of the set of variables that produces the best values of one or more objective functions while complying with multiple constraints. The purpose of optimization has been described as objective function, loss function, or cost function for minimization and utility function or fitness function for maximization [1] [2]. In this paper, it will be referred to as objective function. Simulation

optimization (SO) refers to the optimization of an objective function subject to constraints, both of which can be evaluated through a stochastic simulation/digital model [3]. The term simulation optimization (SO) is an overall term for techniques used to optimize stochastic simulations. Simulation optimization involves the search for those specific settings of the input parameters to a stochastic simulation such that a target objective, which is a function of the simulation output, is either maximized or minimized [3]. Simulation techniques allow for modelling and artificially reproducing complex systems using stochastic distributions [4]. Complex simulation models may require long development times and difficult verification and validation processes and finally, simulation is not an optimization tool on its own [5]. According to [5] large Combinatorial Optimization Problems (COPs) require the use of metaheuristics to conduct an efficient search, where he proposes to combine simulation with metaheuristics to form a new class of optimization algorithms called 'simheuristics'. These algorithms integrate simulation (in any of its variants) into a metaheuristic-driven framework to solve complex stochastic COPs. A metaheuristic is a high-level problem-independent algorithmic framework that provides a set of guidelines or

strategies to develop heuristic optimization algorithms [6]. The JaamSim simulation package used in this industrial case study currently has no optimization analysis capability [7]. It is thus proposed by the author to develop and integrate a metaheuristic genetic algorithm optimization engine with the JaamSim Tray Loader digital model thus enabling the optimization of this industrial case system.

An optimization problem involves searching for an optimal solution(s) x_i from a search space X , which maximize (or minimize) an objective function $f(x)$, while satisfying a set of constraints [8]. The search space X may be composed of discrete variables (e.g., integer, categorical), continuous variables or mixed variables [9]. Metaheuristics are general algorithmic frameworks, often nature-inspired, designed to solve complex optimization problems [10]. Metaheuristics are a growing research area over the last number of years. Metaheuristics are emerging as successful alternatives to more classical approaches also for solving optimization problems that include in their mathematical formulation uncertain, stochastic, and dynamic information [10]. The Greek suffix “meta” used in the word metaheuristic means “beyond, in an upper level”. Thus, metaheuristics are algorithms that combine heuristics (that are usually very problem-specific) in a more general framework. Metaheuristics are strategies that guide the search process. The goal is to efficiently explore the search space to find near-optimal solutions. Techniques which constitute metaheuristic algorithms range from simple local search procedures to complex learning processes [11]. Optimization algorithms attempt to improve solutions in each iteration, seeking to converge toward the optimal solution. After a number of iterations, the search reaches an optimal region of the feasible decision space. The best solution calculated by the algorithm at the time of termination constitutes the optimal solutions of a particular run. **Fig. 1** portrays the process of optimization by Metaheuristic and evolutionary genetic algorithms.

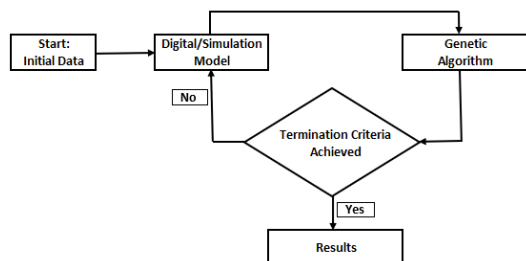


Figure 1. Components of the Optimization System using Simulation and Genetic Algorithms

II. GENETIC ALGORITHMS

1. Genetic Algorithm Overview

Among the meta-heuristic optimization methods, genetic algorithms have gained importance because of its capacity to find sets of optimal solutions [12]. A genetic algorithm (GA) is an 'intelligent' probabilistic search algorithm which simulates the process of evolution by taking a population of solutions and applying genetic operators in each reproduction [13]. Genetic Algorithms (GAs) are adaptive heuristic search algorithms based on the evolutionary ideas of natural selection and genetics. They are a part of evolutionary computing, a rapidly growing area of artificial intelligence. GAs are inspired by Darwin’s theory of evolution – “Survival of the fittest”. Simplicity of operation and power of effect are two of the main attractions of the GA approach [14]. Genetic algorithms are popular as they are relatively easy to implement and are used in several commercial software packages [3]. Genetic algorithms (GA) have been used for the resolution of a wide variety of combinatorial problems, due to the demonstrated success in the results it can achieve [15]. Despite the advantages of genetic algorithms, several parameter inputs are required before using this algorithm. They include waypoint, population size, crossover rate, and mutation rate. The potential GA solution to a problem is an individual which can be represented by the set of parameters. These parameters are just like a gene of a chromosome and can be represented by the string of values in binary form [16]. The fitness value is used to test the degree of goodness of the chromosome for solving a problem that is directly related to the objective value. The operators employed in a GA include selection, crossover, and mutation processes [16] [17].

The performance of the Genetic Algorithm is dependent on these parameter settings [18]. The GA method requires the algorithm to be initialized with a set of randomly generated initial values, which is known as initial population which represents a significant difference with respect to mathematical programming techniques. The initial population is then evaluated to determine which of the individuals have the best characteristics (i.e., the best values for the objective functions), allowing them to pass to the next generation (or iteration). There is a similarity between GA and those that can be observed with the natural evolution concepts. Once the population has been evaluated, the best individuals combine their genetic information between them, and a new generation is obtained. Standard GAs begins with a randomly generated population of possible solutions (individuals). The individual’s fitness is calculated and some of them are selected as parents according to their fitness values. A new population (or generation) of possible solutions (the children’s population) is produced by applying the crossover

operator to the parent population and then applying the mutation operator to their offspring. The fitness value is recalculated for this new population. The iterations involving the replacement of the original generation (old individual) with a new generation (children) is repeated until the termination criteria is achieved. This whole process is shown in **Fig. 2**.

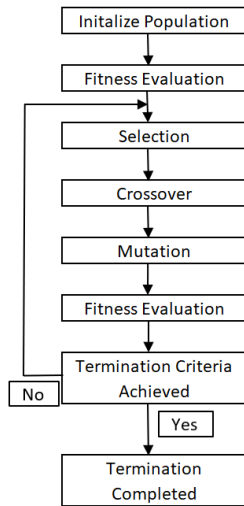


Figure 2. Genetic Algorithm Flowchart [19]

2. Elitism Strategy

A solution with a high fitness value could be replaced by a weaker solution after a crossover or mutation occurs. The process of maintaining good solutions with high fitness after a certain generation cannot be guaranteed. Hence an elitism strategy can be applied in GA to maintain a certain number of the fittest solutions for the next generation. When the next-generation population is obtained after crossover and mutation, these solutions that were maintained by elitism will replace the weaker solutions. The same number of the fittest solutions will replace the weaker solutions and be retained and utilized for the next generation [19] [20]. It has been shown that results obtained by an algorithm which uses elitism is better than the result obtained by an algorithm which doesn't use elitism [21], [22].

3. GA Parameters and Termination Strategy

The size of the population of solutions (M), the number of parents (R), the probability of crossover (PC), the probability of mutation (PM), and the termination criterion are the user defined parameters of the GA. A good choice of the parameters is related to the decision space of a particular problem, and in general the optimal parameter setting for one problem may not perform equally as well for other problems. Consequently, determining a good parameter setting often requires the execution of many time-consuming experiments. A critical factor in implementing a genetic algorithm is how to set the

values for the various parameters. [23] classifies these efforts into two major forms:

1. Parameter tuning. It refers to finding good values for the parameters before the algorithm is run and then keeping these values fixed while the algorithm runs. With this method, typically one parameter is tuned at a time, which may cause some suboptimal choices, since parameters often interact in a complex way with each other. Simultaneous tuning of more parameters, however, leads to an enormous number of experiments.
2. Parameter Control. This method forms an alternative, as it amounts to starting a run with initial parameter values which are then changed during the run.

Selecting the appropriate GA parameters is regularly done based on experience with specific optimization problems. However, a reasonable method for finding suitable values for the GA parameters is to perform sensitivity analysis. This entails choosing a combination of GA parameters and running the GA several times. Other combinations of parameters are chosen, and repeated runs are made with each combination. A comparison of the optimization results obtained may lead to the best set of GA parameters. The author has used Design of Experiments to select the optimum GA parameters for the Tray Loader application.

A termination criterion is required to allow the Genetic Algorithm to end its iterations. Selecting an appropriate termination criterion has an important role on the correct convergence of the algorithm. The number of iterations, the amount of improvement of the objective function between consecutive iterations, and the run time are common termination criteria for the GA.

4. NSGA-II (Non-dominated Sorting Genetic Algorithm II)

The non-dominated sorting algorithm (NSGA), developed in 1994, was one of the first Multi Objective Evolutionary Algorithms (MOEA) [24]. NSGA differs from the standard GA in the way that the selection operator performs, with the crossover and mutation operators remaining the same. The population of solutions is ranked based on its nondomination before selection takes place. Improvements to NSGA were made to tackle issues such as high computational complexity, lack of elitism, need to specify sharing parameter and a technique was added to embed constraints into the optimization algorithm, leading to a new algorithm known as NSGA-II being introduced [25]. According to [26] [27] one of the most widely used MOEA's that has been effective in finding the Pareto optimal solutions is the elitist NSGA-II algorithm. Both the diversity and the convergence abilities of the NSGA-II algorithm have been demonstrated by

[28]. They have also shown the suitability of NSGA-II in producing an acceptable number of optimized design alternatives regarding the problem complexity and in a reasonable timeframe. A detailed review of NSGA-II optimization algorithm in machining operations was presented by [29]. They concluded that NSGA-II as part of Multi Objective Optimization Problem (MOOP) is a popular and reliable algorithm that can be used in optimizing the process parameters of multiple machine performances. Unlike the single objective optimization technique, NSGA-II simultaneously optimizes each objective without being dominated by any other solution [29]. The problem of controlling an air conditioning system using evolutionary algorithms to increase energy-saving while also considering user satisfaction was investigated [30]. They concluded that the NSGA II as an excellent algorithm for solving a multi objective optimization problem. It has also been shown that the multi-objective optimization technique NSGA-II applied to a project was efficient in searching for multiple solutions and was able to find a pareto front after a few iterations during the optimization process [31]. NSGA-II applies an elitist strategy which improves the convergence of an MOEA and avoids the loss of optimal solutions after getting them [32]. It is proposed to use the Elitist NSGA-II and develop a standalone multi objective optimization engine that will run fully integrated with the JaamSim Tray Loader digital model. The workings of the NSGA-II will now be further explained. The flowchart for NSGA-II is shown in Fig. 3.

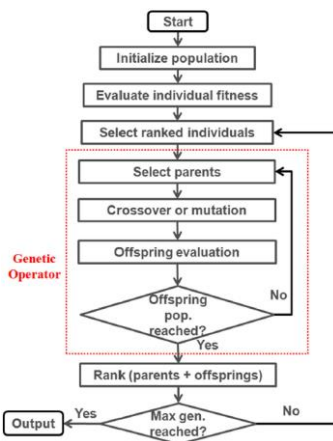


Figure 3. NSGA-II Algorithm Flowchart [31]

In NSGA-II parents and offspring are combined, followed by non-dominated sorting. The fitness of all individuals is assessed and chosen to be parents for the next generation. The NSGA-II Non-dominated sorting and crowding distance sorting, which is depicted in Fig 4 is then completed. P_t is the parent generation and Q_t the offspring that are both merged into R_t . The objective is to obtain a new generation P_{t+1} of the same size as the parent population P_t . Two parameters are estimated for

each individual: the domination count, which provides the information of how many solutions dominate the individual, and a list of the set of solutions that are dominated by the individual. This method splits up all solutions into different fronts. As per Fig. 4, PF_{1-3} are the fronts that are obtained by the sorting process.

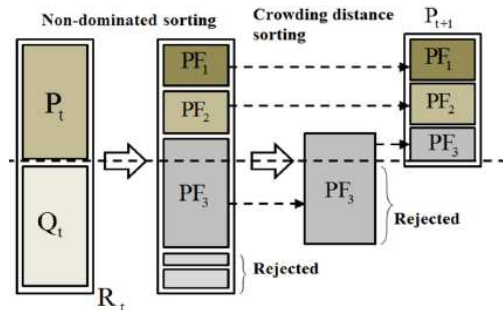


Figure 4. NSGA-II Ranking Procedure [25]

All individuals are compared with each other. The first front will comprise only solutions with a domination count of 0. From there, the algorithm continues going individual by individual through all sets of solutions that have a domination count of 0 to form the first front. The individuals from this 1st front are removed from the list, and the remaining individuals now compared to each other with the 2nd front obtained by selecting individuals with a new domination count of 0. After this process, all the individuals that have a domination count of zero, excluding the first front solutions, will form the second front. The procedure is continued until the last front is obtained as can be seen in Fig. 5.

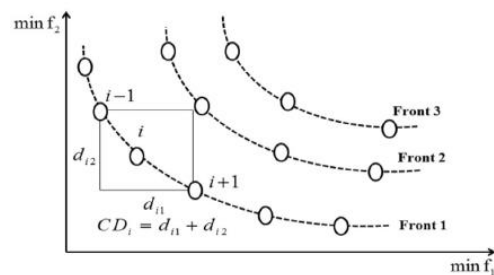


Figure 5. Solution Pareto fronts and Crowding Distance Estimation

From Fig 5 all solutions in PF_1 and PF_2 are taken forward to the new population P_{t+1} . Some solutions from PF_3 are taken forward to P_{t+1} , while the remainder is rejected. The solutions that are taken forward from PF_3 is based on the crowding distance calculation, with the lesser crowded distance individual being chosen to form the total in population P_{t+1} [25]. The crowding distance is a number that determines how closely other solutions are surrounding an individual. Figure 5 shows the calculation of the crowding distance of solution i . The crowding distance is an estimate of the size of the largest cuboid enclosing solution i without including any other solution [30]. The nearest neighbours are used to calculate the average distance

between the closest solutions of the same front. A higher value of crowding distance gives a lesser crowded region and vice versa [25].

III. DEVELOPMENT OF TRAY LOADER DIGITAL MODEL

5. Overview of The Tray Loader Digital Model

A digital model of an industrial system (**Fig. 6**) known as a Tray Loading System was developed using JaamSim software.

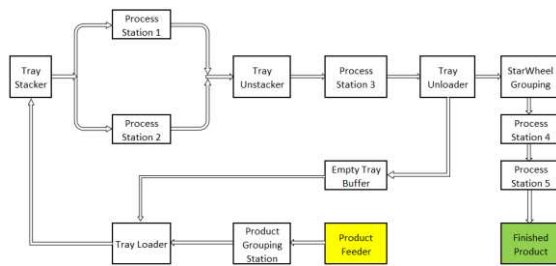


Figure 6. Automated Tray Loading System Industrial Case

This system consists of individual product (p) that arrives from an upstream line to a product feeder at defined arrival times. These are then grouped into multiples of 10. The group of products are then loaded into empty plastic trays that can hold up to 660 parts. Once filled the plastic tray moves at a defined cycle time to a tray stacker. The tray stacker accumulates the filled trays into groups of 30. This group of 30 trays then undergoes a batch process in either Process station 1 or 2 under defined conditions. Upon completion of this batch process, the trays of product leave Process Station 1 or 2, where a tray unstacking operation takes place. Each individual tray of product undergoes a further process step (Process Station 3), again under defined conditions. Once a tray is finished at Process Station 3, the product is removed from the tray at the Tray Unloading station and is then passed to the Star Wheel grouping station, where the product is now grouped into batches of 30. These groups are then passed to Process Station 4 and 5 for the final finishing process. The empty trays from the tray unloading station, are returned to the empty tray buffer and finally back to the tray loader operation, to repeat the overall process. The digital model developed, will simulate this whole operation, considering the following 5 points:

1. Entities (units of Product) per arrival.
2. Service times for process stations, travel times for conveyors
3. Probability distributions for reliability and repair of stations.
4. Conditions for process stations to process and pass product to the next station.

5. Queue size and location.

6. Verification of the Tray Loader Digital Model

A detailed verification process was undertaken on the Tray Loader digital model following the Logical/mathematical verification, program/code verification steps outlined by [33] and the detailed knowledge of the author of the actual tray loading system. All the Tray Loader Objects, Service Times, Steps, Thresholds, Maintenance conditions and Threshold condition logic were all verified and confirmed to be correct to how the actual line operates. A detailed verification checklist was completed on the Tray Loader digital model. As part of the digital model verification process it was important to verify that the product flow into and out of the various simulation objects (as seen from the JaamSim GUI) are identical to what occurs on the tray loader line. This verification process allowed any additions or changes to the simulation logic to be corrected, verified, and visualized immediately. It was through the ongoing and iterative model verification and the testing process during model development, that a realistic model of the actual dynamic interactions was developed and fine-tuned. During this phase of model verification, the weak points of the system were discovered and corrected. It is extremely advantageous to find these early-stage simulation bugs, thus allowing a well-tested and robust system to be developed.

7. Validation of the Tray Loader Digital Model

The approach taken for developing the Tray Loader digital model followed the steps described by [34]. Step 5 of this approach deals with confirming that the programmed model is valid. The model is run using the standard basic settings from the actual tray loader system. The simulation model output data for the system was compared with the comparable output data collected from the actual system. This is called results validation. If the results are consistent with how the system should operate, then the simulation model is said to have face validity. Sensitivity analyses is performed on the programmed model to see which factors have the greatest impact on the performance measures and, thus, must be modelled carefully [34]. According to [35], validation is concerned with determining whether the conceptual digital model (as opposed to the computer program) is an accurate representation of the system under study. [35] outlines the following three (3) steps to validate a simulation model.

1. Obtaining real-world data from the actual system.

2. Tests for comparing simulated and real data (namely graphical, Schruben-Turing or t tests).
3. Sensitivity analysis (using statistical design of experiments with associated regression analysis).

The above approach was used to validate the Tray Loader digital model, see section 3 for more detail. Actual Tray Loader system data was collected from the historian database for all the relevant process stations used in the digital model. The data collected included input feed rate, yield, throughput and uptime per minute for each process station. Excel macros were then developed to calculate the equipment reliability metrics namely: Mean Time Between Failures (MTBF) and Mean Time to Repair (MTTR) for each of the process stations using the uptime/minute data. The Input feed rate, yield, output data and the MTBF/MTTR for each process station was analysed, outliers removed, and distributions determined along with the distribution parameters. Minitab is used to analyse all the data obtained. Minitab is a statistical analysis software that assists in the analysis of data collected from any process and provides a simple, effective way to input the data, manipulate that data and statistically analyse it.

IV. DEVELOPMENT OF THE NSGA-II OPTIMIZATION ENGINE

A closed loop digital model and optimization engine is proposed by the author as shown in Fig. 7. An NSGA-II optimization engine is integrated with the Tray Loader JaamSim digital model and the following four (4) elements being executed automatically until an optimized solution is obtained:

1. Digital Model inputs parameters updated.
2. Simulation runs executed and monitored
3. Digital Model outputs collected.
4. Optimization analysis completed and new parameter settings recommended.

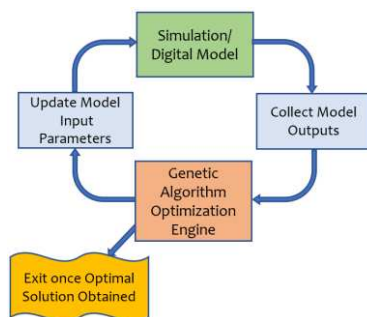


Figure 7. Closed Loop Optimization Engine

The overall optimization system developed by the author allows the user to specify the objectives to be optimized from an excel file. Table 1 shows an example of two (2) objectives to be minimized along

with two (2) objectives to be maximized (station throughputs). This file is used to configure the optimization problem along with the associated objectives to be either maximised or minimised.

Table 1. Optimization Objectives

Processes	Objective
Max Trays Used	Minimize
P_Feeder	Maximize
Process4	Maximize
P_Feeder Util	Minimize

Another excel file is set-up to store all the entities/workstations names along with their associated base parameter values, see Table 2 for a sample of some configuration settings for the Tray Loader Simulation model.

Table 2. Simulation Model and Optimization Parameters

Entity_Name	Parameter	Value	Optim_Space
P_Feeder	InterArrivalTime	0.90 s	-10 to 10
P_Feeder	Downtime Duration (MTTR)	2.00 Min	-10 to 10
Tray_Pack	Downtime Duration (MTTR)	2.00 Min	-10 to 10
Process4	ServiceTime	2.3 s	-10 to 10
Process5	ServiceTime	2.3 s	-10 to 10
P_Feeder_Yield	Weibull Location	0.5573	-10 to 10
Empty_Tray_Stacker	Capacity	90	-15 to 15

The settings in the Optim_Space column are used by the NSGA-II optimization engine. As an example, referring to Table 2, JaamSim is configured with an entity generator called P_Feeder. The base InterArrivalTime for this generator is 0.90sec. When performing an optimization analysis, the InterArrivalTime for this entity can be changed within a space of 0.90 sec $\pm 10\%$ in increments of 1%. Likewise, the Tray Loader JaamSim model uses a resource called Empty_Tray_Stacker, with a base setting of 90 units. The optimization space for this parameter is 90 $\pm 15\%$ in increments of 1%. The user can select which parameters, the base setting for that parameter and if required the optimization space for that parameter to be used by NSGA-II.

Reviewing the optimization space for the 7 factors in Table 2, there is in excess of 2.6 Billion combinations of different factor settings that the Tray Loader line can be operated to. It is impossible to run all of those combinations using the Tray Loader digital model, hence the need to use optimization approaches to determine a particular setting for each of the 7 factors that results in an optimum solution to the required objective(s). Python code was developed that integrates the excel input configuration files with both the JaamSim Tray Loader digital model and the NSGA-II optimization engine. The overall structure and Python code that was written to integrate the NSGA-II and the JaamSim Tray Loader digital model to form an optimization system followed a modular format. This modular format followed the ten (10) rules and two (2) best practices for code development

highlighted by [36]. The overall system architecture is shown in **Fig. 8**. This architecture gives a high-level overview of how the optimization system was developed with the main optimization system being controlled by the module called *invoke_simw* (see purple box in **Fig. 8**). The main function module called *invoke_simw* then calls other blocks (red boxes in **Fig. 8**) forming the main spine of the Tray Loader Optimization system. All the function modules are written using the python programming language. The four (4) main blocks of the system include:

1. Main controlling function module called *invoke_simw*
2. Input Data Pre-processing function block that calls several sub function modules.
3. Overall GA and JaamSim Optimization Loop function block calling several sub function modules.
4. Output data file post processing block calling several sub function modules.

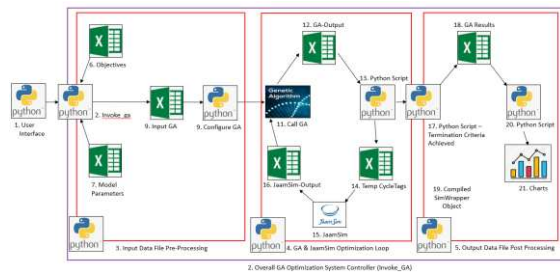


Figure 8. OPTIM-GA Program and Data Structure

Tray Loader digital model parameters are passed to the ‘*invoke_simw*’ function. The *invoke_simw* function (**Fig. 8**), then schedules the calling of all the various functions and methods required to execute all the tasks in the three (3) red boxes. When all input data pre-processing is completed, the NSGA-II and JaamSim Optimization loop (**Fig. 8**) is activated where simulation runs are completed using the tray loader digital model. Output results from each simulation run is then analyzed by NSGA-II optimization engine and any associated changes to the digital model input parameters based on the requirements of the objective function are then made. This process is repeated until the termination criteria is achieved thus producing an optimal solution. The tray loader termination criteria is reviewed in section 8 below.

Once NSGA-II optimization has terminated the program returns to the calling function ‘*invoke_simw*’. At this point the function ‘Output Data File Post Processing’ **Fig. 8** is called. This block of code prepares the results from the optimization study for review and graphing. The data is also saved to a csv file to allow the user to further analyze the data with statistical packages (eg: Minitab ©) to support any decisions in relation to possible design changes to the tray loading system. A significant number of Python libraries associated

with optimization have been developed recently, however, only a few of them support optimization of multiple objectives at a time [37]. As such, pymoo (python multi-objective optimization) which is a library of multi-objective optimization tools was developed in Python [37]. There are several different algorithm implementations in ‘pymoo’ examples include GA and NSGA-II to name a few. These NSGA-II pymoo library of optimization routines were used in the development of the overall Tray Loader NSGA-II optimization system. A plug-in library called pymoo, Ver 0.5.0 was then installed into the Thonny IDE to enable multi objective optimization in Python (**Fig. 9**).

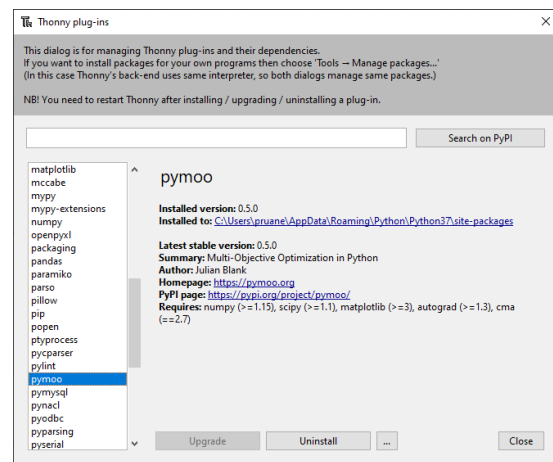


Figure 9. Pymoo Library in the Thonny IDE

A list of the additional Python libraries installed into the Thonny IDE are given in **Table 3**. These libraries are required to allow the developed python code for the Tray Loader optimization to run without errors.

Table 3. Python Libraries installed into the Thonny IDE

Number	Library Name	Installed Version	Summary Description
1	beautifulsoup4	4.10.0	Screen-scraping library
2	bitstring	3.1.7	Simple construction, analysis and modification of binary data.
3	bs4	0.0.1	Dummy package for Beautiful Soup
4	KDE-diffusion	1.0.3	Kernel density estimation via diffusion in 1d and 2d
5	lxml	4.6.4	Powerful and Pythonic XML processing library combining libxml2/libxslt with the ElementTree API.
6	matplotlib	3.3.3	Python plotting package
7	numpy	1.19.5	NumPy is the fundamental package for array computing with Python.
8	openpyxl	3.0.9	A Python library to read/write Excel 2010 xlsx/xlsm files
9	pandas	1.2.0	Powerful data structures for data analysis, time series, and statistics
10	pymoo	0.5.0	Multi-Objective Optimization in Python
11	thonny	3.3.13	Python IDE for beginners
12	scipy	1.7.1	SciPy: Scientific Library for Python
13	seaborn	0.11.2	seaborn: statistical data visualization

8. NSGA-II Termination Criteria for the Tray Loader Optimization Problem

Whenever an optimization algorithm is executed, it needs to be determined at each iteration whether the optimization run shall be continued or not. Many different ways exist of how to decide when an optimization run should be terminated. Running the algorithm not long enough can lead to unsatisfactory results and running it too long might waste function evaluations, time and thus computational resources. Pymoo have developed several termination criterion for both single and multi-objective optimization. The Tray Loader termination criteria uses the standard ‘Termination’ function which was imported by python from pymoo. Actual code is given below:

```
from pymoo.core.termination import
Termination
```

According to [38] the most interesting stopping criterion is to use objective space change to decide whether to terminate the algorithm. This termination criteria uses a simple and efficient procedure to determine whether to stop the optimization or not. This termination procedure is called ‘MultiObjectiveSpaceToleranceTermination’, and is imported from pymoo as given by the actual code below:

```
from pymoo.util.termination.f_tol
import
MultiObjectiveSpaceToleranceTermin
ation
```

This termination procedure ‘MultiObjectiveSpaceToleranceTermination’ is then configured with various termination parameters and assigned to the ‘termination’ attribute with python code as given below:

```
# NSGA-II Tray Loader termination
criteria

termination =
MultiObjectiveSpaceToleranceTermin
ation(tol=0.0025, n_last= min(30,
n_max_gen), nth_gen= min(5,
patience), n_max_gen= n_max_gen,
n_max_evals=None)
```

The five (5) termination parameters [38] above are described as follows:

1. `tol` = This is the average threshold tolerance in the objective space. If the value is below this bound (0.25% from above), the algorithm is terminated.

2. `n_last` = To make the termination criterion more robust, this parameter specifies the last n generations to review and then takes the maximum from this number of generations.
3. `nth_gen` = Defines whenever the termination criterion is calculated by default, or every n th generation. In the example above, `nth_gen` is the minimum of 5 or the patience value.
4. `n_max_gen` = Furthermore, the number of generations executed by the algorithm can be used for termination. For some optimization problems, the termination criterion might not be reached, thus, an upper bound for generations can be defined to stop in this case.
5. `n_max_evals` = Lastly, the number of function evaluations can be used for termination. In the example above, this is not used as can be seen when this variable is set to **None**.

9. NSGA-II parameter tuning for the Tray Loader Application

The key to a successful implementation of Genetic Algorithms primarily depends on the efficient crossover and mutation search operators to guide the system toward a global optimum [39]. The values of GA parameters greatly determine whether the GA will find a near-optimum solution and whether it will find such a solution efficiently in a timely manner. Choosing the right parameter values can be a time-consuming task where the computer specifications can play a significant factor in how long it takes to obtain both the GA optimum parameters and determining the optimum solution to the problem itself [23]. According to [40], GAs are not easy to use because they require parameter tunings in order to achieve the desirable solutions. The task of tuning GA parameters has been proven to be far from trivial due to the complex interactions among the parameters. In the research carried out by [41], parameter setting for MOOP using evolutionary algorithms (MOEAs) is crucial for finding the best performance of the algorithm. These parameters are very sensitive in driving the algorithms to the best performance and finding the good results. Design of experiments (DoE) methods offer practical approaches to tune the parameters effectively [42]. It has been shown that the internal parameters of NSGA-II can be tuned using the Design of Experiments (DoE) procedure to enhance the quality of the results for the synthesis optimization of a four-bar mechanism [43]. The six (6) operating parameters of the NSGA-II algorithm which need to be set for the Tray Loader optimization application are as follows:

1. Population size.
2. # of Offspring.

3. Crossover Probability
4. Crossover Distribution Index
5. Mutation Probability
6. Mutation Distribution Index

These parameters affect the capability of the algorithm to achieve the optimum objective results and computing time to reach these results. According to [44] population size can be decided by experience and usually between 50 and 160. If the population size is too small, then it can be difficult to get an optimum solution, whereas, if it's too large then the convergence time can be long. A recommended range of parameter settings is given in **Table 4** to achieve optimum GA performance [41], [42], [44].

Table 4. GA Parameter Settings

GA Parameter Name	Identifier	Range
Population Size	P	30 - 100
Crossover Probability	P _C	0.2 - 0.8
Mutation Probability	P _M	0.001 - 0.1

The mutation distribution index (η_m) and the crossover distribution index (η_c) are typically set in the range of 10 – 40 [45]. A large crossover distribution index (η_c) gives a higher probability for creating near parent solutions and a small crossover distribution index (η_c) allows distant solutions to be selected as children solutions [46]. The parameters with the associated levels for each parameter that are used in the NSGA-II algorithm are given in **Table 5**.

Table 5. NSGA-II Parameters and Levels

Factor #	Factor Name	Symbol	# of Levels	High Setting	Centre Setting	Low Setting
1	Population Size	P	2	100	65	30
2	# of Offspring	P _O	2	0.80	0.55	0.30
3	Crossover Probability	P _C	2	0.80	0.55	0.30
4	Crossover Distribution Index	η_c	2	40	25	10
5	Mutation Probability	P _M	2	0.10	0.05	0.01
6	Mutation Distribution Index	η_m	2	40	25	10

A ½ fractional DoE was chosen for tuning the Tray Loader NSGA-II optimization parameters as the resolution provided was sufficient to analyse the data. A total of 33 runs is required for the experiment (32 ½ fraction runs and 1 centre level run). Each experiment was run for a maximum of 30 generations based on previous optimization experiments carried out by the author during the development of this optimization system. Increasing the number of generations, significantly increases the time required to run each experiment. The max *P_Feeder* output and max *Process4* output was recorded across the total population for each of the thirty (30) generations. Analysis of Variance (ANOVA) and response optimization of the NSGA-II parameters is completed using Minitab in order to maximise both the *P_Feeder* and *Process4* outputs. The results are shown in **Table 6**.

Table 6. Tray Loader NSGA-II Parameter Optimization

Solution						
Solution	Population Size	No of Offspring	Crossover Probability	Crossover Distribution Index	Mutation Probability	Mutation Distribution Index
1	1	1	-1	-1	-1	1
Process P_Feeder						
4 Result Result Composite						
Solution	Fit	Fit	Fit	Desirability		
1	441228	461766	0.870449			

Based on this analysis, the Tray Loader NSGA-II optimization system is configured with the parameter values as shown in **Table 7**.

Table 7. Tray Loader NSGA-II Parameter Values

Parameter #	ParameterName	Symbol	Setting Level	Setting Value
1	Population Size	P	1	100
2	# of Offspring	P _O	1	0.80
3	Crossover Probability	P _C	-1	0.30
4	Crossover Distribution Index	η_c	-1	10
5	Mutation Probability	P _M	-1	0.01
6	Mutation Distribution Index	η_m	1	40

V. RESULTS FROM THE TRAY LOADER DIGITAL MODEL AND NSGA-II OPTIMIZATION ENGINE

All simulation/optimization runs were completed using a HP ZBook Firefly 15 G7 2Z4F7UC laptop running an Intel(R) Core(TM) i7-10810U CPU @ 1.61 GHz processor and 64GB of RAM. The single objective optimization run (Maximize *P_Feeder* output) was executed 10 times as recommended by [47] [48]. The results of the 10-run experiment is given in Table 8.

Table 8. Simulation Model and Optimization Parameters

Run Number	Total # of Generations	# of Generations to Best Sol'n	Total Execution Time (Mins)	Time to Optimal Sol'n (Mins)	P_Feeder Mean	P_Feeder Max	P_Feeder Min	P_Feeder STD
1	50	21	481	202	459,575	462,391	457,550	1,322
2	50	50	470	470	459,345	461,870	458,117	1,076
3	50	40	544	435	458,955	462,741	457,305	1,367
4	50	50	460	460	454,725	461,899	424,117	1,182
5	25	8	262	84	457,000	461,870	455,286	1,539
6	45	26	473	273	458,702	462,685	456,726	1,432
7	30	14	278	130	457,656	462,741	455,715	1,519
8	50	35	518	363	458,204	461,654	456,585	1,230
9	45	26	442	255	459,158	462,391	457,623	1,050
10	25	9	236	85	457,868	462,741	455,828	1,593
Average of All Runs	42	28	416	276	458,119	462,298	453,485	1,331

As can be seen from **Table 8**, the *P_Feeder* Mean, Max, Min and Standard Deviation is calculated across the 50 generations for each run using the NSGA-II optimization. The average *P_Feeder* (max) across the 10 runs using NSGA-II was 462,298 units. The overall *P_Feeder* maximum output across the 10 runs using NSGA-II optimization was 462,741 achieved on runs 3, 7 and 10. Run #1 was analysed in additional detail, as the results of this particular run produced results that were close to the overall average of the 10 runs completed. Analyzing the data collected from Run

#1, the P_Feeder maximum, minimum, average and standard deviation is calculated for each of the 50 generations and plotted using Minitab©. **Fig. 10** shows how all the individual solutions within the population of 100 solutions for each of the 50 generations are converging closer to the P_Feeder maximum value of 462,391 which was achieved on generation #22.

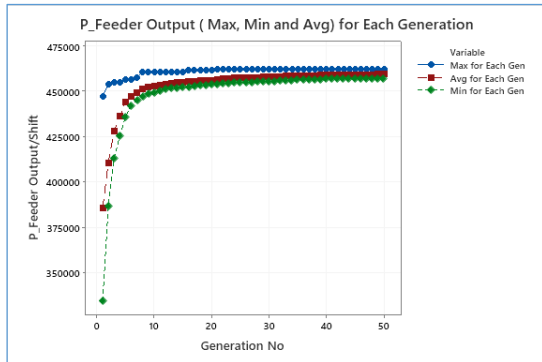


Figure 10. NSGA-II Optimization of P_Feeder Output/Shift

The maximum P_Feeder output remained unchanged for the remaining 28 generations of the experiment. The standard deviation of the P_Feeder output within the population of 100 solutions for each generation is plotted and can be seen in **Fig. 11**.

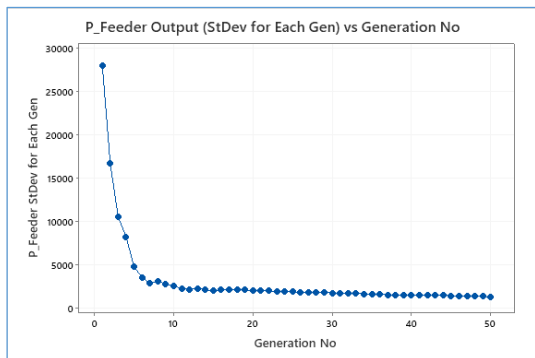


Figure 11. NSGA-II optimization of P_Feeder Standard Deviation

Fig. 11 shows that as the solutions are generated for each generation, the spread is reducing indicating that all of the solutions are progressively getting closer to the optimum P_Feeder max value of 462,391 and the optimization procedure can be terminated. It can be seen from **Fig. 10** and **11** that the NSGA-II algorithm has converged after approximately 22 generations, at which point the fitness value function (max P_Feeder Output) was unchanged and the standard deviation of P_Feeder output of all the solutions within each generation decreasing slightly. To reduce the optimization computation time, the maximum number of generations could be reduced from 50 to approx. 30. This value of 30 was selected (greater than 22), with the aim of avoiding an early termination of the

algorithm before the max P_Feeder output was obtained. The solution developed for run #7 (**Table 8**) with an overall P_Feeder max of 462,741 units using the Tray Loader JaamSim digital model and the NSGA-II optimization engine is shown in **Table 9**.

Table 9. Tray Loader Optimized Digital Model Parameters

Digital Model Used	Optimization Engine Type	P_Feeder (Max)	Optimized Tray Loader Digital Model Parameters						
			P_Feeder Inter Arrival Time	P_Feeder Yield (Location)	P_Feeder Down Time_Duration	Process Down Time (Sec)	Tray_Pack Down Time_Duration	Process Down Time (Sec)	Empty_Tray Stack Capacity Count
Tray Loader JaamSim	NSGA-II	462,741	81	0.613	2.04	2.41	2.18	2.50	94

A two (2) objective optimization problem (maximize the P_Feeder and minimize the Empty Tray Buffer capacity) was designed and tested using the tray loader digital model and NSGA-II optimization engine. As with the single objective optimization problem, the same Tray Loader simulation model, model parameters and optimization parameter space was used for this study (See **Table 2**), with results given in **Table 10**.

Table 10. Two Objective Optimization problem of Tray Loader System using NSGA-II Optimization

Run Number	NSGA-II Optimization (Population size = 100, Analysis of last Generation)							
	Total # of Generations	# of Generations to Best Sol'n	Total Execution Time (Mins)	Time to Optimal Sol'n (Mins)	P_Feeder Min	P_Feeder Max	Avg_Trays_Used (Min)	Avg_Trays_Used (Max)
1	50	31	453	281	304,943	461,707	61	84
2	40	22	355	195	311,928	461,592	63	84
3	30	11	432	158	308,997	461,650	63	84
4	50	34	434	295	308,967	461,917	62	81
5	50	17	438	149	303,533	462,391	61	84
6	35	18	301	155	316,285	460,309	64	85
7	45	28	435	271	315,424	462,741	63	81
8	35	13	350	130	305,053	460,750	63	95
9	50	38	464	353	312,300	461,273	63	84
10	35	11	355	112	305,756	460,683	62	84
Average of All Runs	42	22	402	210	309,319	461,501	63	85

A pareto front is a set of nondominated solutions, being chosen as optimal, if no objective can be improved without sacrificing at least one other objective [8]. The pareto front is an excellent visualization to show the interaction of each objective has on the other. A pareto front (*Empty Tray Buffer Capacity vs P_Feeder Output/Shift*) was generated using all the data gathered from the 2 objective SimWrapper Optimization runs. See **Fig. 12** for the pareto front.

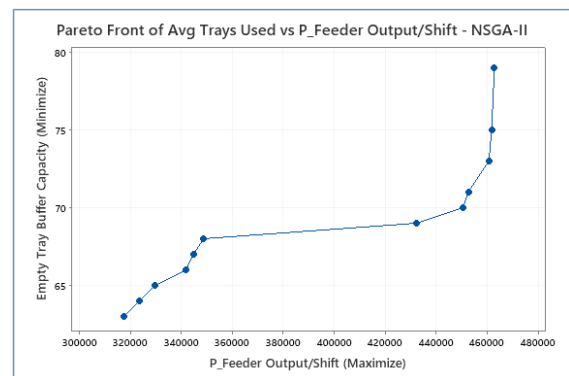


Figure 12. Tray Loader 2 Objective Optimization Pareto Front.

As can be seen from Fig. 12, the optimum solution is where the Empty Tray count (Buffer Capacity) is approx. 79 trays, thus producing a stable P_{Feeder} output of ~ 462,741. Increasing the Empty Tray buffer beyond 79 trays, has no impact on the P_{Feeder} output/shift. Since the optimization problem is to minimize Empty Tray Buffer and maximize P_{Feeder} output, the factor setting providing the solution of 79 trays and P_{Feeder} output of 462,741 is selected.

VI. CONCLUSION

As manufacturing capital equipment is expensive, it is necessary that the equipment once in operation is reliable and delivers to the business plan targets. Simulation along with an optimization system is an invaluable tool to confirm that an automated manufacturing line can produce to the required business objectives before and after it goes into operation. Implementing the actual changes to equipment to improve reliability can be both time consuming and expensive. Simulation in conjunction with optimization can be used to verify these improvements before the equipment is modified. These technologies form the basis of an overall digital manufacturing system that enables the optimization of a manufacturing line during the line design stage or when the line is put into operation. The use of this technology gives a deeper understanding of what can occur on the manufacturing line when it is running. A simulation model when combined with optimization engine, can be used to identify problems before they occur and aid in the selection of optimum parameters to run the line before it is fully designed or built. Digital model and optimization technologies supports other Industry 4.0 technologies such as predictive maintenance, OEE improvement, waste reduction, improve batch changeover times and to improve product quality [49]. It allows for efficient

REFERENCES

- [1] W. Erwin Diewert, "The New Palgrave Dictionary of Economics," Palgrave Macmillan UK, 27 April 2017. [Online]. https://link.springer.com/referenceworkentry/10.1057/978-1-349-95121-5_659-2.
- [2] S. Boyd and L. Vandenberghe, Convex Optimization, vol. 7, Cambridge University Press, 2009.
- [3] S. Amaran, N. Sahinidis, B. Sharda and S. Bury, "Simulation Optimization: A Review of Algorithms and Applications," *Annals of Operations Research*, vol. 240, pp. 351-380, 2016. <https://doi.org/10.1007/s10479-015-2019-x>
- [4] R. Nance and R. Sargent, "Perspectives on the Evolution of Simulation," *Operations Research*, vol. 50, no. 1, pp. 161-172, 2002. <https://doi.org/10.1287/opre.50.1.161.17790>
- [5] A. Juan, J. Faulin, S. Grasman, M. Rabe and G. Figueira, "A review of Simheuristics: Extending Metaheuristics to deal with Stochastic Combinatorial Optimization Problems," *Operations Research Perspectives*, vol. 2, pp. 62-72, 2015. <https://doi.org/10.1016/j.orp.2015.03.001>
- [6] K. Sörensen and F. Glover, "Metaheuristics," In: Gass, S.I., Fu, M.C. (eds) *Encyclopedia of Operations Research and Management Science*. Springer, Boston, MA (2013) pp.

design and development, linking 3D models with simulation and emulation of equipment control code. In addition, having a digital model enables virtual line analysis, removing the physical restraints of expert engineers having to be on your location [50]. The author has demonstrated how the development of digital model can be validated and subsequently used as part of an optimization system which is then used for the study of equipment design, maintenance and reliability of an automated production line in the medical devices industry.

ACKNOWLEDGEMENT

The authors express their sincere gratitude to Johnson & Johnson Vision Care and Technological University of the Shannon for providing me with the tools, time and support necessary to allow me to pursue this very important research in the field of simulation and optimization of advanced automated manufacturing equipment.

AUTHOR CONTRIBUTIONS

Patrick Ruane: Conceptualization, development, Experimentation, Analysis, Writing and Editing.

Patrick Walsh: Supervision & Review.

John Cosgrove: Supervision & Review.

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

Patrick Ruane <http://orcid.org/0000-0002-6685-1663>

- 960-970.
https://doi.org/10.1007/978-1-4419-1153-7_1167
- [7] King, D.H, and H.S Harrison. 2013. "Open Source Simulation Software "JaamSim"." Proceedings of the 2013 Winter Simulation Conference. Washington, DC, USA. <https://doi.org/10.1109/WSC.2013.6721593>
- [8] N. Gunantara and Q. Ai, "A review of multi-objective optimization: Methods and its applications.," *Cogent Engineering*, vol. 5, no. 1, pp. 1 - 16, 2018. <https://doi.org/10.1080/23311916.2018.1502242>
- [9] J. Pelamatti, L. Brevault, M. Balesdent, E. Talbi and Y. Guerin, "Efficient global optimization of constrained mixed variable problems.," *Journal of Global Optimization*, vol. 73, no. 3, pp. 583-613, 2019. <https://doi.org/10.1007/s10898-018-0715-1>
- [10] L. Bianchi, M. Dorigo and L. Gambardella, "A Survey on Metaheuristics for Stochastic Combinatorial Optimization," 2009. <https://doi.org/10.1007/s11047-008-9098-4>
- [11] C. Blum and A. Roli, "Metaheuristics in Combinatorial Optimization Overview and Conceptual Comparison," *ACM Computing Surveys*, vol. 35, no. 3, pp. 268-308, 2003. <https://doi.org/10.1145/937503.937505>
- [12] F. Castro, C. Gutierrez-Antonio, A. Briones-Ramirez and J. Hernandez, "Genetic Algorithms: A tool for Optimizing Intensified Distillation Sequences," *Genetic Algorithms: Advances in Research and Applications*, pp. 1-17, 2017.
- [13] P. Chu and J. Beasley, "A Genetic Algorithm for the Generalised Assignment Problem," *Computers & Operations Research*, vol. 24, no. 1, pp. 17-23, 1997. [https://doi.org/10.1016/S0305-0548\(96\)00032-9](https://doi.org/10.1016/S0305-0548(96)00032-9)
- [14] D. Goldberg and J. Holland, "Genetic Algorithms and Machine Learning," *Machine Learning*, vol. 3, no. 2-3, pp. 95-99, 1988. <https://doi.org/10.1023/A:1022602019183>
- [15] M. Dalle Mura and G. Dini, "A Multi-Objective Software Tool for Manual Assembly Line Balancing using a Genetic Algorithm," *CIRP Journal of Manufacturing Science and Technology*, vol. 19, pp. 72-83, 2017. <http://dx.doi.org/10.1016/j.cirpj.2017.06.002>
- [16] R. Haupt and S. Haupt, *Practical Genetic Algorithms*, New York, NY, USA: John Wiley & Sons, 2004. <https://doi.org/10.1002/0471671746>
- [17] M. Hamza, H. Yap and I. A. Choudhury, "Genetic Algorithm and Particle Swarm Optimization Based Cascade Interval Type 2 Fuzzy PD Controller for Rotary Inverted Pendulum System.," *Mathematical Problems in Engineering*, vol. 2015, pp. 1 - 15, 2015. <https://doi.org/10.1155/2015/695965>
- [18] P. Rajendran and K. Yit Yok, "The Optimization of a Genetic Algorithm for Unmanned Aerial Vehicle Path Planning," *Genetic Algorithms: Advances in Research and Applications*, pp. 19 - 32, 2017.
- [19] K. Kok, P. Rajendran, R. Rainis, I. Wan and M. Wan Mohd, "Investigation on selection schemes and population sizes for genetic algorithm in unmanned aerial vehicle path planning.," *International Symposium on Technology Management & Emerging Technologies (ISTMET)*, pp. 6 - 10, 2015. <https://doi.org/10.1109/ISTMET.2015.7358990>
- [20] K. Yit Kok, P. Rajendran, R. R and W. Mohd Muhiyuddin Wan Ibrahim, "Investigation on Selection Schemes and Population Sizes for Genetic Algorithm in Unmanned Aerial Vehicle Path Planning," in *2015 International Symposium on Technology Management and Emerging Technologies (ISTMET)*, Langkawi, Malaysia, 2015. <https://doi.org/10.1109/ISTMET.2015.7358990>
- [21] C. Grosan and M. Oltean, "The Role of Elitism in Multiobjective Optimization with Evolutionary Algorithms," *Acta Universitatis Apulensis* 5 (2003) pp. 83-90.
- [22] G. Guariso and M. Sangiorgio, "Improving the Performance of Multiobjective Genetic Algorithms: An Elitism-Based Approach," *Information*, 11 (12) pp. 1 - 14, 2020. <https://doi.org/10.3390/info11120587>
- [23] A. Eiben, R. Hinterding and Z. Michalewicz, "Parameter Control in Evolutionary Algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 2, pp. 124-141, 1999. <https://doi.org/10.1109/4235.771166>
- [24] N. Srinivas and K. Deb, "Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms," *Evolutionary*

- Computation, p. 22 1– 248, 1994.
<https://doi.org/10.1162/evco.1994.2.3.221>
- [25] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, “A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182 - 197, 07 August 2002.
<https://doi.org/10.1109/4235.996017>
- [26] T. Goel, R. Vaidyanathan, R. Haftka, W. Shyy, N. Queipo and K. Tucker, “Response Surface Approximation of Pareto Optimal Front in Multi-Objective Optimization,” *Computer Methods in Applied Mechanics and Engineering*, vol. 196, no. 4, pp. 879 - 893, 2007.
<https://doi.org/10.1016/j.cma.2006.07.010>
- [27] C. Leung and H. Lau, “Simulation-based optimization for material handling systems in manufacturing and distribution industries,” *Wireless Networks*, p. 4839 – 4860, 2020.
<https://doi.org/10.1007/s11276-018-1894-x>
- [28] N. Mahammed, M. B. S, O. A and M. Fahsi, “Evolutionary Business Process Optimization using a Multiple-Criteria Decision Analysis method,” in *International Conference on Computer, Information and Telecommunication Systems (CITS)*, 2018.
<https://doi.org/10.1109/CITS.2018.8440166>
- [29] Y. Yusoff, M. Salihin Ngadiman and A. Mohd Zain, “Overview of NSGA-II for Optimizing Machining Process Parameters,” *Procedia Engineering 15*, p. 3978 – 3983, 2011.
<https://doi.org/10.1016/j.proeng.2011.08.745>
- [30] A. Ruiz, S. Martínez, J. Rocha, J. Villanueva, J. Menchaca, M. Berrones, M. Flores and A. Pineda, “Assessing a Multi-Objective Genetic Algorithm with a Simulated Environment for Energy-Saving of Air Conditioning Systems with User Preferences,” *Symmetry 2021*, vol. 13, no. 2, 20 February 2021.
<https://doi.org/10.3390/sym13020344>
- [31] Y. Chang, Z. Bouzarkouna and D. Devegowda, “Multi Objective Optimization for Rapid and Robust Optimal Oil Field Development Under Geological Uncertainty,” *Computational Geosciences*, vol. 19, p. 933 – 950, 2015.
<https://doi.org/10.1007/s10596-015-9507-6>
- [32] H. Kumar and S. Yadav, “Hybrid NSGA-II Based Decision Making in Fuzzy Multi Objective Reliability Optimization Problem,” *SN Applied Sciences*, 2019.
<https://doi.org/10.1007/s42452-019-1512-2>
- [33] P. K. Davis, “Generalizing Concepts and Methods of Verification, Validation and Accreditation (VV&A) for Military Simulations,” *National Defense Research Institute*, pp. 5-6, 1992.
- [34] A. M. Law, "How to Build Valid and Credible Simulation Models," 2019 Winter Simulation Conference (WSC), 2019, pp. 1402-1414.
<https://doi.org/10.1109/WSC40007.2019.9004789>
- [35] J. P. Kleijnen, “Theory and Methodology of Verification and validation of simulation models,” *European Journal of Operational Research*, vol. 82, pp. 145-162, 1995.
- [36] H. Hunter-Zinck, A. de Siqueira, V. Vásquez, R. Barnes and C. Martinez, “Ten Simple Rules on Writing Clean and Reliable Open-Source Scientific Software,” *PLOS Computational Biology*, pp. 1 - 9, 2021.
<https://doi.org/10.1371/journal.pcbi.1009481>
- [37] J. Blank and K. Deb, “Pymoo - Multi-objective Optimization in Python,” *IEEE Access*, vol. 8, pp. 89497 - 89509, 2020.
<https://doi.org/10.1109/ACCESS.2020.2990567>
- [38] Blank, J, and K Deb. 2020. “A Running Performance Metric and Termination Criterion for Evaluating Multi and Many Objective Optimization Algorithms.” 2020 IEEE Congress on Evolutionary Computation (CEC) pp. 1 - 9.
<https://doi.org/10.1109/CEC48606.2020.9185546>
- [39] Lim, S.M, A.B Sultan, N Sulaiman, A Mustapha, and K.Y Leong. 2017. “Crossover and Mutation Operators of Genetic Algorithms.” *International Journal of Machine Learning and Computing* 7 (1).
<https://doi.org/10.18178/ijmlc.2017.7.1.611>
- [40] Duc Tran, Khoa. 2005. “Elitist non-dominated sorting GA-II (NSGA-II) as a parameter-less multi-objective genetic algorithm.” *Proceedings. IEEE Southeast Conference. Ft. Lauderdale, FL, USA.* pp. 359 - 367.
<https://doi.org/10.1109/SECON.2005.1423273>
- [41] Samsuri, S, R Ahmad, M Zakaria, and M Zain. 2019. “Parameter Tuning for Comparing Multi-Objective Evolutionary Algorithms Applied to System Identification Problems.” *Proc. of the 2019 IEEE 6th International Conference on Smart Instrumentation, Measurement and Applications.* Kuala

- Lumpur, Malaysia.
<https://doi.org/10.1109/ICSIMA47653.2019.9057333>
- [42] Arin, A, G Rabadi, and R Unal. 2011. "Comparative studies on design of experiments for tuning parameters in a genetic algorithm for a scheduling problem." *International Journal of Experimental Design and Process Optimisation* 102-124. <https://doi.org/10.1504/IJEDPO.2011.040262>
- [43] Badduri, J, R.A Srivatsan, Kumar G.S, and S Bandyopadhyay. 2012. "Coupler-Curve Synthesis of a Planar Four-Bar Mechanism Using NSGA-II." *Asia-Pacific Conference on Simulated Evolution and Learning*. 460-469. https://doi.org/10.1007/978-3-642-34859-4_46
- [44] Cao, Z, and Z Zhang. 2010. "Parameter Settings of Genetic Algorithm Based on Multi-Factor Analysis of Variance." 2010 Fourth International Conference on Genetic and Evolutionary Computing. Shenzhen, China. 305 - 307. <https://doi.org/10.1109/ICGEC.2010.82>
- [45] Deb, K. 2011. "Multi-Objective Optimization Using Evolutionary Algorithms." Department of Mechanical Engineering, Indian Institute of Technology Kanpur, Kanpur, PIN 208016, India, 1 - 24. Accessed January 10, 2022. <https://www.egr.msu.edu/~kdeb/papers/k2011003.pdf>
- [46] Deb, K, and H Beyer. 2001. "Self Adaptive Genetic Algorithms with Simulated Binary Crossover." *Evolutionary Computation* 9 (2): 197 - 221. <https://doi.org/10.1162/106365601750190406>
- [47] M. Jeong, J. H. Choi and B. H. Koh, "Performance evaluation of modified genetic and swarm-based optimization algorithms," *Structural Control and Health Monitoring*, p. 878–889, 2013. <https://doi.org/10.1002/stc.507>
- [48] J. Shen and Y. Zhu, "Chance-Constrained Model for Uncertain Job Shop Scheduling Problem," *Soft Computing - A Fusion of Foundations, Methodologies & Applications.*, vol. 20, no. 6, pp. 2383-2391, June 2016. <https://doi.org/10.1007/s00500-015-1647-z>
- [49] G. Shao, S. Jain, C. Laroque, L. H. Lee, P. Lendermann and O. Rose, "Digital Twin for Smart Manufacturing: The Simulation Aspect," 2019 Winter Simulation Conference (WSC), 2019, pp. 2085-2098. <https://doi.org/10.1109/WSC40007.2019.9004659>
- [50] Q. Qi, F. Tao, T. Hu, N. Anwer, A. Liu, Y. Wei and L. Wang, "Enabling Technologies and Tools for Digital Twin," *Journal of Manufacturing Systems*, vol. 58, pp. 3-21, 2021.



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.