The Long-Range Macroeconomic Effects of Sars-Covid-19 Pandemic in Hungary: a Conceptual Framework and Methodology – Focusing Approach

Éva Szendrő, Zoltán Lakner

Hungarian University of Agriculture and Life Sciences
Páter Károly u. 1, H-2100 Gödöllő, Hungary
e-mails: szendro.eva.monika@uni-mate.hu; lakner.zoltan.karoly@uni-mate.hu

Abstract: The COVID pandemic has caused an unprecedented socio-economic situation in the world, short-term consequences of which are well documented. At the same time, the long-term effects of the crisis are much lesser studied and quantified. The goal of the current paper is to demonstrate a potential way of analysis for quantification of economic consequences of the pandemic based on a combination of econometric, health statistics and health economic data. The direct losses in the long-range, caused by COVID related death up to the end of January 2022, compared to the predictable economic development trajectory of Hungary were estimated as 1.2-1.4 milliard USD, discounted to the current value. Results of analysis highlight the importance of health care system development in the maintenance of economic dynamism as well as health-related intervention programs.

Keywords:Cobb-Douglas function; disease burden; simulation; system dynamics modelling

1 Introduction

It is well documented that one of the biggest shocks to the modern world economy has been the pandemic, caused by the virus strain "*severe acute respiratory syndrome coronavirus 2*" [1] (SARS-CoV-19, hereinafter: COVID). Its short-term consequences are well treated on the scale of the global economy [2] [3] [4], of different countries [5] as well as in various branches (e.g. [6] [7]) and practically in all spheres of socio-economic activities [8]. At the same time, we have a rather vague concept of its long-range macroeconomic consequences. The purpose of the current paper is twofold: (1) presentation a general, simple, but transparent and relatively easily applicable conceptual model, based on a combination of macro-, demographic- and health economic data to determine the long-range consequences of the Covid crisis; (2) application of this approach to the Hungarian economy to calculate the direct, long-range consequences of deaths, caused by the pandemic.

2 Methodology

2.1. Conceptual Framework

The backbone of our approach is the application of the Solow-Swan model of economic development. This framework is a very simple model of long-run economic growth of Solow, 1956 [9] and Swan, 1956 [10], based on quantification of the mutual, dynamic relationship of production factors and value-added in a given national economy. The conceptual framework of the model is depicted in Fig. 1.

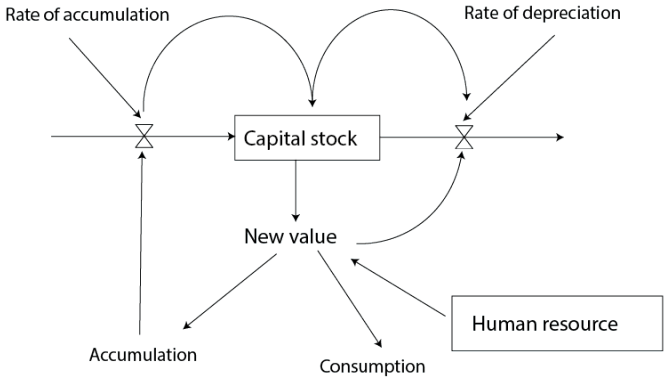


Figure 1
The conceptual model of economic development, based on the theory of Solow and Swan

Results, presented in the current paper, are based on a rather complex workflow, the basic idea of which has been originally developed for quantification of burden of chronic non-communicable diseases by the World Health Organisation (WHO) [11]. Our workflow consisted of four main steps. These were as follows:

1) In the first phase of research, we have determined the stochastic relationship between the two basic production fact///ors: capital and labour and the new value-added, approximated by the GDP. Traditionally, this relation is based on Cobb-Douglas function [12] [13] [14]. However, this method and its theoretical foundations are sometimes questioned [15], it is widely applied to describe the lows, governing economic growth [16] [17]. The classic form of this equation is a constant elasticity of substitution (CES) function:

$$y = \gamma\left(\delta x_1^{-\rho} + (1-\delta)x_2^{-\rho}\right)^{\frac{-\nu}{\rho}}$$

Where y is the value-added of the national economy, x_1 and x_2 are the production factors (in our case the capital and the labour) γ,ρ,δ,ν are parameters. The function (1) is not linear in parameters, and cannot be linearised on the base of traditional methods, that's why we had to apply

different non-linear least-square optimisation algorithms. For this purpose, we have used the micEconCES r–package [18], which offers a wide range of algorithms for fitting and estimation of parameters.

The latest publications on CES functions [19] take into consideration the effects of quality changes in human resources. These approaches are mainly based on the equation, proposed by Mincer [20] to measure the economic effect of schooling. For the sake of completeness, we present the changes in quality of the living labour, but we had been focussing on quantitative changes in human resources because the inclusion of estimations on the future level of human resource qualification could make the model a rather speculative one.

Based on these pieces of information we had been able to construct a theoretical model of Hungarian economic development for the next decades, without taking into the consequences of COVID–related additional burden of disease. This model will be called hereinafter as the baseline model.

2) In the next part of the research, we have determined the additional burden, caused by the COVID pandemic. Based on the conceptual model, depicted in Fig. 1, we have assumed that the pandemic created economic losses will affect long-range, beyond the immediate economic backdrop. (1) This will be a consequence of COVID-related deaths, decreasing the human workforce; (2) The premature deaths of COVID decrease the population, and this process generates structural changes in health care costs. As a consequence of these two processes, there are changes in production factors in the national economy, which based on circular causality will lead to further changes in macroeconomic performance. This trajectory of economic development has been called as a post-COVID model. This model is a dynamic one, too, which considers the yearly changes in the number of different cohorts from year to year in a way, depicted in Fig. 2.

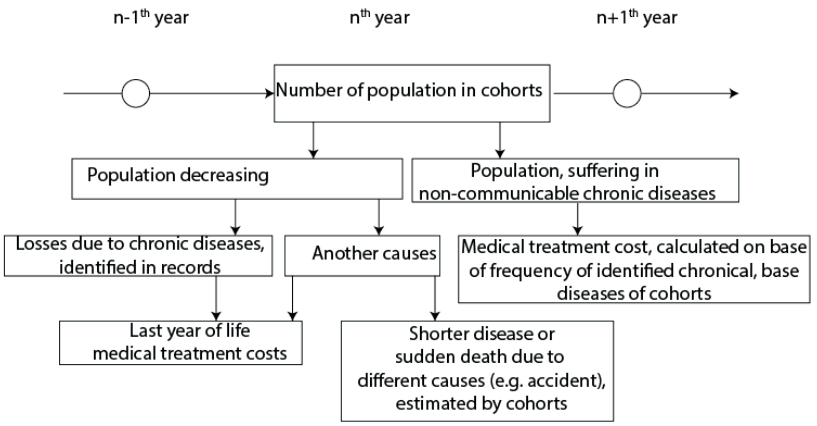


Figure 2
Model for medical cost estimation

3) The comparison of baseline and the post-COVID model offers a possibility to quantify the long-range effects of the COVID pandemic. To compare the different monetary values, we have applied the Net Present Value (NPV) concept Gaspars-Wielloch [21]. According to this approach, the present value of a future cash flow can be calculated by the formula (2).

$$NPV = \sum_{t=0}^{t=n} \frac{CF_t}{(1+i)^t}$$

where CF_t is the cash flow at the t_{th} time period and i is the discount rate.

4) The different parameters of the model are based on estimations, with a relatively high level of uncertainty. That's why we have applied simulations techniques to quantify the sensitivity of the model.

2.2. Operationalisation of the Research Plan and Data Management

2.2.1 Estimation of the Baseline Scenario

Data for the determination of Cobb-Douglas function have been downloaded from Penn World Table [22] [23]. This, internationally widely recognised and applied [24] [25] database contains a wide range of parameters, from which we have used the indices as follows: Real GDP at constant 2017 national prices, converted to 2017 USD, number of persons engaged, capital stock at constant 2017 national prices, converted to 2017 USD.

The general demographic processes in Hungary for the next decades have been determined on the basis of UN population forecast database [26]. We have supposed that the average working-age will begin at age 20 and end at the age of 65. We calculated with a 70% participation in the work of active population. This is in line with the current share. According to the data of the Central Statistical Office of Hungary [27] the average age of people, entering into a pension in 2019 has been 63 years, but in the case of younger generations, the age limit of work will be 65 years. We have calculated with this value, supposing that each person, who is capable of working is willing to labour as a full-time worker. At the same time, we have supposed, that the people in pension will not work in any form (which is – of course – a considerable simplification). We have not had enough information on the severity of different diseases, that's why we have estimated that the work intensity of individuals with diseases will be lower by20 per cent.

The average depreciation rate has been 4.5792%, determined as an average of estimations for 2015-2019 years in Penn database. The gross capital formation ratio has been estimated on the basis of the World bank Database [28]. This ratio has

been 25.483% of the GDP in the average of the 2016-2020 years. We have supposed that this ratio will be standard for the next few decades.

In our calculations, we have supposed that the GDP of Hungary in 2023 will return to the level of 2019. In this way, we have not taken into consideration the short-term economic consequences of COVID.

2.2.2 Estimation of Parameters to Determine the Economic Losses

Data for the determination of prevalence and general causalities of different diseases have been downloaded from the Global Burden of Disease Collaborative Network database [29]. Currently, this database can be considered as a „gold standard" of disease-prevalence related calculations [30] [31] all over the world.

The most critical part of the study has been the determination of the health condition of people, who died in the COVID. The data collection on COVID-related mortality and morbidity data is not standardised on the international level, that's why the comparison of original (raw) data is not possible [32]. The competent Hungarian authority regularly issued the data of deceased individuals on a specific website [33], but the indication of different diseases has not been in line with any standards. The cause of death has been indicated often rather generally or according to the jargon, applied in the actual hospital. A detailed critique of this system has been written by Ferenci [34]. His heroic work [35] has been an essential contribution for data cleaning: e.g. the name of high blood pressure as a disease has been written in 24(!) variants in the original database.

The dataset, cleaned and created by Ferenci [36] has been the stepping stone for further analysis. In the next phase, we have determined the economic consequences of death due to COVID. For this purpose, by the built-in random number generator of Excel we have chosen a sample of 200 dead people from causalities of pandemic below 60 years, and different five-years age brackets from 60 to 85 years. We have analysed their health condition one by one, including two colleagues with MSc-level health-care qualifications. One of the authors and these specialists independently classified the health status of the persons in the samples and determined the most life-threatening disease, if there has been different diseases indicated.

On the basis of GBD dataset, we have estimated expected life-years with the given disease. In case of casualties, where no disease has been indicated, we have supposed a mortality rate, which has been calculated for the given cohorts, by GBD database. If there has not been given any information on the base disease of the deceased, we have projected the relative share of diagnosed diseases on this group. The deceased people, who did not have any known basic disease have been dealt with separately.

Costs of treatment of different diseases have been collected from the peer-reviewed literature, on the basis of heuristic research of Pubmed and Web of Science databases.

According to the European Cardiovascular Disease Statistics [37] the prevalence of cardiovascular diseases in Hungary has been 1.22 million people, the total cost of treatment was 1.511 milliard €, the cost per capita was 153 €/year.

In the case of cancer-related diseases, the direct economic costs of cancer treatment in Hungary has been 393 million € [38], the estimated prevalence is 225 thousand, that's why the costs were estimated as 3500 €/year.

Neural diseases cover a very wide range of diseases. From this follows that the costs of treatment are estimated on the basis of the article of Kovács et al. (2020) [39] as 1000 €/year/patient.

The most important component of the disease group is diabetes. There are different estimations of diabetes treatment costs [40], we have applied the cost review of Stegauer et al. (2020) [41]. According to their systematic review, the median of diabetes specific direct costs according to different studies is approximately 500 € annual excess cost for patients with type 2 diabetes. The chronic kidney disease costs show considerable differences as a function of the severity of the disease. The lowest costs of these treatments are in the range of 14-80 thousand USD/year.

The comparison of health care costs in different countries is an extremely difficult task. This problem is well treated in the literature [42] [43]. For simplicity, we have applied the comprehensive study of Koechline et al., 2017 [44].

Diernberger et. al [45] estimate the health treatment of last year's costs in Great Brittain as 10000 Pound Sterling. According to the calculations of [44] price relations of health care systems in Hungary and Great Brittain are in 35:110 ratio, that's why in case of Hungary we have calculated by 3100 GB Pounds cost in last year of life, which is equal to 3700 € for the last year of life.

For calculation of NPV we have applied a conservative 6% time preference (discount) rate.

3 Results and Discussion

In the first phase, we have determined the Cobb-Douglas function of the Hungarian economy on the basis of data from 1970 to 2019. If we consider the dynamics of the most important indicators, obviously there are rather contradictory tendencies (Fig. 3). The GDP had been increasing rather rapidly between 1970 and the middle of the eighties of the last century. At the end of the centrally planned economy, the GDP witnessed stagnation. The system transformation caused a considerable backdrop, but this decrease in value-added creation has been counterbalanced by rapid development, which had been fuelled by international conjuncture.

The economic performance has been considerably increasing after the years of the world economic crisis in the last years of the first decade of the new millennium. The capital stock continuously increased, however, the investment cycles remained a characteristic feature of the Hungarian economy (Fig. 4). It is worth highlighting the increase of intellectual capital. The working population considerably decreased after the system changed. In the last decade, there is a rapid increase in it.

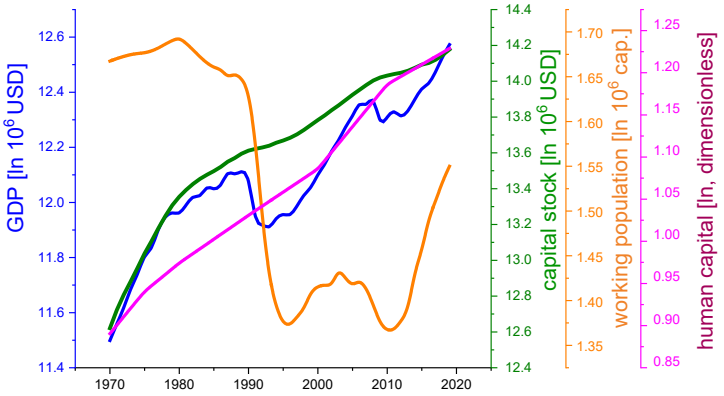


Figure 3

Changes of GDP and factors of production in Hungary, between 1970 and 2019

Source: own compilation, based on [23]

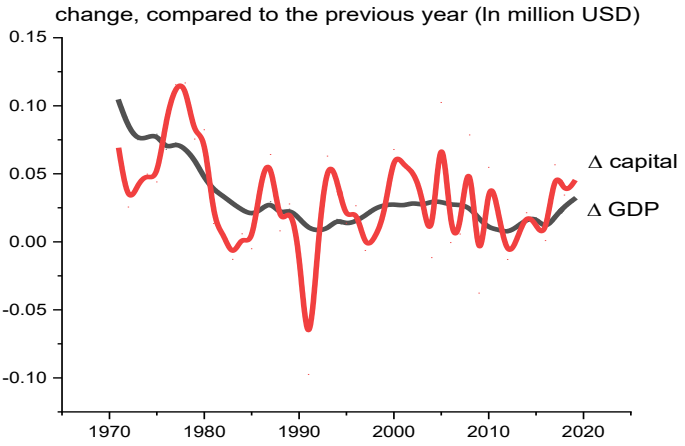


Figure 4

The roller coaster of Hungarian investment cycles

Source: own compilation, based on [23]

Analysing the results of fitting of the CES function (Table 1), it is obvious that the R squared values in some cases (e.g. Kmenta approximation, Hicks neutral technological change) are not interpretable. In the case of numerous methods, the

fitting of the function is rather high, but the elasticity of substitution values show a considerable difference.

Table 1
Determination of Cobb-Douglas function parameters by different algorithms

Algorithms	γ	δ	ρ	υ	Multi. R^2	Elasticity of substitution
Kmenta approximation	5.94*	-0.89*	-0.50**	1.36**	-495.44	0.66
Conjugate gradients	0.81	0.42	-1.89	1.23**	0.98	-1.12
Conjugate gradients with increased level of tolerance	2.95	1.00**	-0.02	0.53*	0.71	1.02
Newton-type method	1.62	1.73	-0.35	0.56**	0.89	1.55
Broyden-Fletcher-Goldfarb-Shanno algorithm	2.75	0.00	-5.01	1.27**	0.99	0.05
Nelder-Mead global optimisation	1.16	0.11	-3.75	1.14**	0.98	1.79
Simulated Annealing	0.67	0.73	-1.84	1.17**	0.98	-1.17
Differntial evolution	4.95	0.99	-0.75	0.33	0.22	10.01
Byrd, Lu, Nocedal and Zhu algorithm	0.52	0.95	-0.37	1.22	28.26	1.59
Quasi-Newton Port routine	0.30**	0.84**	-1.00	1.43**	0.98	NA
Hicks neutral technological change (lambda=9.42 6E-3)	3.0E-8	6.29E-1	1.89	1.04	-3.19	0.34
Grid search algorithm	0.52	0.96**	-0.30	1.22	0.98	1.42
Two-stage grid search	0.60	0.77**	-1.21	1.23	0.985	-4.78

Legend: * significance at 95%, ** significance at 99%

We have applied the parameters, obtained by application of Newton-type method for further work. Obviously (Fig. 5), the function, which has been determined, are suitable to describe the long-range development trajectory of the Hungarian economy in the last fifty years. There is a wide choice of different functions, but the difference between the various function is marginal.

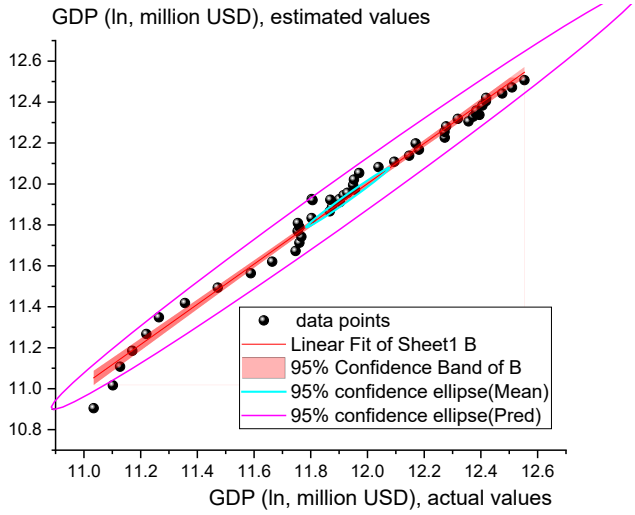


Figure 5
Fitting of the estimated CES function

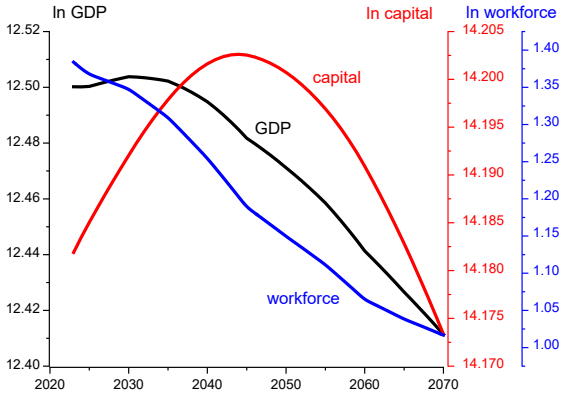


Figure 6
Characteristic parameters of the baseline model

The results of the baseline model (Fig. 6) show a relatively constant decrease in the GDP and the workforce. The relatively high level of capital accumulation rate will increase the capital stock up to the middle of forties, but later on, the demand of workforce will cause a decreased capital stock, and the superposition of decreasing of the workforce and the capital stock will cause a further decreasing of the GDP.

The analysis of the share of the relative share of different diseases in the case of deceased people shows highlights some characteristic features. Obviously (Fig. 7), the share of various diseases does not show very large differences according to cohorts. As a tendency, it can be determined the dominating role of cardiovascular diseases at all age groups. Neurological diseases play an increasing role in the case of elder age brackets.

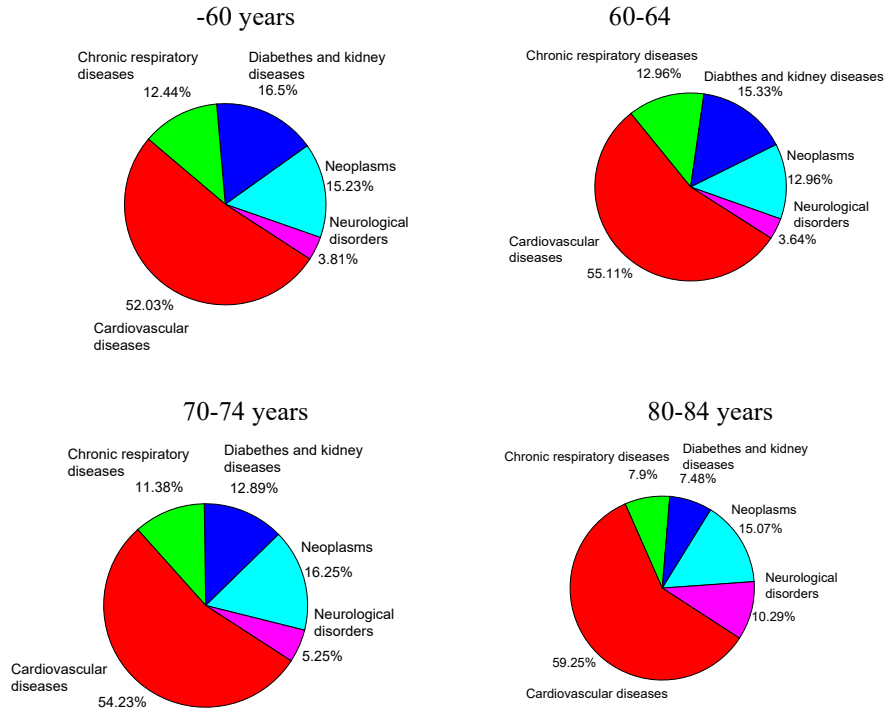


Figure 7

Distribution of known, disease groups by age-brackets

The analysis of the dynamic model to determine the long-range consequences of the pandemics offers two main lessons:

(1) the premature death of working-age members will generate long consequences, the value of which will be between 80-200 million USD in different years. This is equivalent (in current prices) with 18.9-63 milliard HUF. This former is the same order of magnitude, as the budget, allocated for the running of the Ministry of Justice in the Hungarian Budget [1.] for 2022 (19.7 milliard HUF).

(2) If we take into consideration the consequences of re-allocation of healthcare costs, the losses will be lower, especially in the next decades.

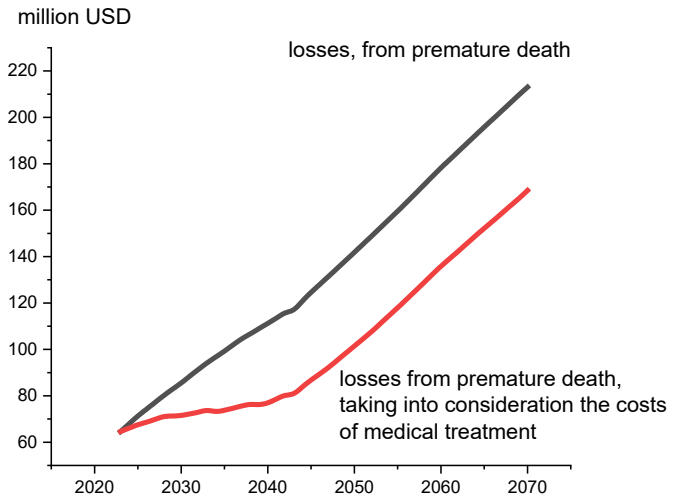


Figure 8

Losses in different years as a consequence of premature death, with and without taking into consideration the medical treatment costs

As we have emphasised earlier, there is a high-level of uncertainty in our calculations, that's why we have analysed the stability of values by simulation. Obviously (Fig. 9) results of simulation of NPV value show that the present value of losses of COVID disease will be between 1.2-1.4 milliard USD. This is double the sum, allocated for higher educational purposes in the Hungarian budget.

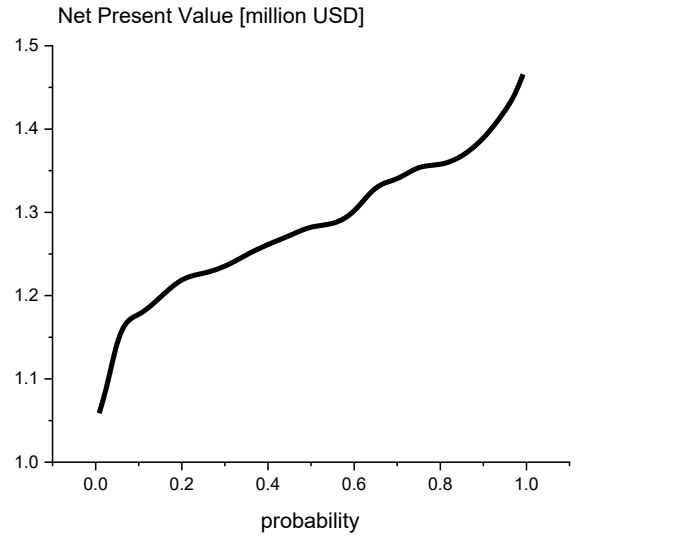


Figure 9

The cumulative distribution function of NPV value of losses, caused by the pandemics

As a summary, it can be stated that the direct, long-range economic consequences of COVID crises are considerable and their adverse effect will be felt in long-time, even if their yearly value is lesser than one tenth per cent of the Hungarian GDP, the net present value of the losses, calculated by 6% discount rate are approximately 1.2-1.6 milliard USD. This fact highlights the importance of the health care system because we have seen that even a moderate decrease of the working population can cause considerable losses. This fact does not counterbalance by lower health care costs, due to premature death of elder generations. Notwithsanding of lack of pieces of information, it can be seen, that the econometric modelling can be a suitable tool for the evaluating and forecasting of the consequences of the pandemics.

4 Limitations

This paper should be considered rather as an interim report, than a final, comprehensive summary of the economic consequences of the COVID crisis, because there is a lack of evidence on the end of COVID pandemics yet. On the approximate date of the end of pandemics, there are different, rather contradictories estimations [47].

The Cobb-Douglas function seems to be an appropriate approach to describe the long-range relation between the capital, but there is room to further sophistication of the model, including other parameters.

The current article has been focusing just on the fatal consequences of the COVID. There is an increasing quantity of proof on long-range adverse consequences of COVID [47], but the time is too short to quantify these. As a consequence of the COVID-caused overburden of the health case system, numerous screenings and not-essential medical interventions were postponed. This fact increases the long-range adverse consequences of the COVID pandemic [48] [49] as well as the adverse psychical consequences Searfini et al.

The effect of the COVID pandemic has modified practically all spheres of socio-economic life. That's why the economic consequences are much deeper and wider as we have presented.

References

[1] B. A. Taha, Y. Al Mashhadany, M. H. H. Mokhtar, M. S. Dzulkefly, B Zan, N. Arsad, An analysis review of detection coronavirus disease 2019 (COVID-19) based on biosensor application. Sensors, Vol. 20, No. 23, 2020, p. 6764

[2] S. Maital, E. Barzani, The global economic impact of COVID-19: A summary of research. Samuel Neaman Institute for National Policy Research, Vol. 2020, No. 2020, pp. 1-12

[3] K. Das, R. L. Behera, B. Paital, Socio-economic impact of COVID-19. COVID-19 in the Environment, pp. 153-190, Elsevier, Amsterdam https://doi.org/10.1016/B978-0-323-90272-4.00014-2

[4] V. Gupta, K. Santosh, R. Arora, T. Ciano, K. S. K. Mohan, Socioeconomic impact due to COVID-19: An empirical assessment. Information Processing & Management, Vol. 59, No. 2, 2022, p. 102810

[5] Ö. Açikgöz, A. Günay, The early impact of the Covid-19 pandemic on the global and Turkish economy. Turkish journal of medical sciences, Vol. 50, No. SI-1, 2020, pp. 520-526

[6] S. M. Jowitt, COVID-19 and the global mining industry. SEG Discovery, Vol. No. 122, 2020, pp. 33-41

[7] Z. Nakat, C. Bou-Mitri, COVID-19 and the food industry: Readiness assessment. Food control, Vol. 121, No. 2021, pp. 1076-1081

[8] M. Batool, H. Ghulam, M. A. Hayat, M. Z. Naeem, A. Ejaz, Z. A. Imran, C. Spulbar, R. Birau T. H. Gorun, How COVID-19 has shaken the sharing economy? An analysis using Google trends data. Economic Research-Ekonomska Istraživanja, Vol. 34, No. 1, 2021, pp. 2374-2386

[9] R. M. Solow, A contribution to the theory of economic growth. The quarterly journal of economics, Vol. 70, No. 1, 1956, pp. 65-94

[10] T. W. Swan, Economic growth and capital accumulation. Economic record, Vol. 32, No. 2, 1956, pp. 334-361

[11] D. Abegunde, A. Stanciole, An estimation of the economic impact of chronic noncommunicable diseases in selected countries. World Health Organization, Department of Chronic Diseases and Health Promotion, Geneva, Vol. 2006, No. 2006

[12] A. Zellner, J. Kmenta J. Dreze, Specification and estimation of Cobb-Douglas production function models. Econometrica: Journal of the Econometric Society, Vol. 5, No. 4, 1966, pp. 784-795

[13] P. H. Douglas, The Cobb-Douglas production function once again: its history, its testing, and some new empirical values. Journal of Political Economy, Vol. 84, No. 5, 1976, pp. 903-915

[14] P. J. Lloyd, The origins of the von Thunen-Mill-Pareto-Wicksell-Cobb-Douglas function. History of Political Economy, Vol. 33, No. 1, 2001, pp. 1-19

[15] S. Gechert, T. Havranek, Z. Irsova D. Kolcunova, Death to the Cobb-Douglas production function? A quantitative survey of the capital-labor substitution elasticity Leibniz Information Centre for Economics, Kiel, Hamburg, html. Vol. No. 2019, https://ideas.repec.org/p/pra/mprapa/95949 [last accessed: 02.02.2022]

[16] D. Fullerton, C. L. Ta, Environmental policy on the back of an envelope: A Cobb-Douglas model is not just a teaching tool. Energy Economics, Vol. 84, No. 2019, pp. 1044-1047

[17] R. Munguía J. Davalos S. Urzua, Estimation of the Solow-Cobb-Douglas economic growth model with a Kalman filter: An observability-based approach. Heliyon, Vol. 5, No. 6, 2019, p. e01959

[18] A. Henningsen G. Henningsen, G. Literáti, Econometric estimation of the" Constant Elasticity of Substitution" function in R: the micEconCES package. Handbook of Research Methods and Applications in Empirical Microeconomics, Edward Elgar Publishing, London, https://doi.org/10.4337/9781788976480, eISBN: 9781788976480

[19] D. E. Bloom S. Chen M. Kuhn M. E. McGovern L. Oxley K. Prettner, The economic burden of chronic diseases: estimates and projections for China, Japan, and South Korea. The Journal of the Economics of Ageing, Vol. 17, No. 2020, p. 100163

[20] J. Mincer, Human capital and economic growth. Economics of education review, Vol. 3, No. 3, 1984, pp. 195-205

[21] H. Gaspars-Wieloch, Project net present value estimation under uncertainty. Central European Journal of Operations Research, Vol. 27, No. 1, 2019, pp. 179-197

[22] R. C. Feenstra, R. InklaarM. P. Timmer, The next generation of the Penn World Table. American economic review, Vol. 105, No. 10, 2015, pp. 3150-3182

[23] https://www.rug.nl/ggdc/productivity/pwt/?lang=en (last accessed: 02.02.2022)

[24] K. Wu M. Wu, Positive Influence of Regulated Human Capital Accumulation on Economic Growth: A Theoretical Model. Economics, Vol. 11, No. 1, 2022, pp. 1-8

[25] A. P. Andrews O. Ijose, The Lewis Growth Rule for Developing Economies. The Journal of Developing Areas, Vol. 56, No. 2, 2022, pp. 253-273

[26] https://population.un.org/wpp/ (last accessed: 02.02.2022)

[27] https://www.ksh.hu/ffi/4-16.html (last accessed: 02.02.2022)

[28] https://data.worldbank.org/indicator/NE.GDI.TOTL.ZS (last accessed: 02.02.2022)

[29] Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Reference Life Table. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2021, http://ghdx.healthdata.org/record/ihme-data/global-burden-disease-study-2019-gbd-2019-reference-life-table (last accessed: 12.12.2022)

[30] I. Iheanacho, S. Zhang, D. King, M. Rizzo A. S. Ismaila, Economic burden of chronic obstructive pulmonary disease (COPD): a systematic literature review. International Journal of Chronic Obstructive Pulmonary Disease, Vol. 15, No. 2020, pp. 439-365

[31] C. J. Murray, C. Abbafati, K. M. Abbas, M. Abbasi, M. Abbasi-Kangevari, F. Abd-Allah, M. Abdollahi, P. Abedi, A. AbediH. Abolhassani, Five insights from the global burden of disease study 2019, The Lancet, Vol. 396, No. 10258, 2020, pp. 1135-1159

[32] J. A. Helliwell, W. S. Bolton, J. R. Burke, J. P. Tiernan, D. G. JayneS. J. Chapman, Global academic response to COVID-19: Cross-sectional study. Learned Publishing, Vol. 33, No. 4, 2020, pp. 385-393

[33] https://koronavirus.gov.hu/elhunytak (last accessed: 12.12.2022)

[34] https://github.com/tamas-ferenci/COVID19MagyarEpi (last accessed: 12.12.2022)

[35] T. Ferenci https://research.physcon.uni-obuda.hu/COVID19MagyarEpi/ (Real-time pandemy of Covid Epidemy in Hungary, [in Hungarian]) (last accessed: 12.12.2022)

[36] https://github.com/tamas-ferenci/C19MortalityLineListingHUN/commit/ece1e5dc39397f6bf62c3b5ecb4c5463f8ae5753 (last accessed: 12.12.2022)

[37] E. Wilkins, L. Wilson, K. Wickramasinghe, P. Bhatnagar, J. Leal, R. Luengo-Fernandez, R. Burns, M. Rayner N. Townsend, European cardiovascular disease statistics, European Heart Network, Brussels 2017, https://ehnheart.org/images/CVD-statistics-report-August-2017.pdf (last accessed: 12.12.2022)

[38] R. Luengo-Fernandez, J. Leal, A. Gray R. Sullivan, Economic burden of cancer across the European Union: a population-based cost analysis. The lancet oncology, Vol. 14, No. 12, 2013, pp. 1165-1174

[39] G. Kovács, T. Almási, A. Millier, M. Toumi, M. Horváth, K. Kóczián, Á. Götze, Z. KalóA. T. Zemplényi, Direct healthcare cost of schizophrenia–European overview. European Psychiatry, Vol. 48, No. 1, 2018, pp. 79-92

[40] Z. FézerL. Kovács, The economic impact of diabetes. 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), Vol. No. 2017, pp. 77-82

[41] C. Stegbauer, C. Falivena, A. Moreno, A. Hentschel, M. Rosenmöller, T. Heise, J. SzecsenyiF. Schliess, Costs and its drivers for diabetes mellitus type 2 patients in France and Germany: a systematic review of economic studies. BMC health services research, Vol. 20, No. 1, 2020, pp. 1-12

[42] Á. Csomós, M. Janecskó D. Edbrooke, Comparative costing analysis of intensive care services between Hungary and United Kingdom. Intensive care medicine, Vol. 31, No. 9, 2005, pp. 1280-1283

[43] J. Xue, J. Hinkle, M.-G. Reeves, L. L. Zheng, V. Natarajan, S. Vyas, R. U. Oli, M. Oliva, R. Kaplan A. Milstein, A cost comparison of cataract surgeries in three countries—United States, India, and Nepal. NEJM Catalyst Innovations in Care Delivery, Vol. 2, No. 9, 2021, pp. 189-195

[44] F. Koechlin, P. Konijn, L. Lorenzoni P. Schreyer, Comparing hospitals and health prices and volumes across countries: a new approach. Social Indicators Research, Vol. 131, No. 1, 2017, pp. 43-64

[45] K. Diernberger, X. Luta, J. Bowden, M. Fallon, J. Droney, E. Lemmon, E. Gray, J. MartiP. Hall, Healthcare use and costs in the last year of life: a national population data linkage study. BMJ Supportive & Palliative Care, Vol. 17, No. 2, 2021, pp. 369-387

[46] 2021. évi XC. törvény Magyarország 2022 évi központi költségvetéséről [Law on Central Budget of Hungary for 2022], Magyar Közlöny, No. 120 pp. 5102-5296

[47] C. WangJ. Han, Will the COVID-19 pandemic end with the Delta and Omicron variants? Environmental Chemistry Letters Vol. 35, No. 1, 2022, pp. 1-11

[48] A. Rivas-Vazquez, G. Rey, A. Quintana A. A. Rivas-Vazquez, Assessment and Management of Long COVID. Journal of Health Service Psychology, Vol. No. 2022, pp. 1-10

[49] V. Phe, G. Karsenty, G. Robert, X. Gamé J.-N. Cornu, Widespread postponement of functional urology cases during the COVID-19 pandemic: rationale, potential pitfalls, and future consequences. Vol. 78, No. 1, 2020, pp. 4-5

Cyber-Security Threats Origins and their Analysis

Maroš Čergeť, Ján Hudec

Faculty of Informatics and Information Technologies,
Slovak University of Technology in Bratislava,
Ilkovicova 2, 842 16 Bratislava, Slovakia
xcerget@stuba.sk, jan.hudec@stuba.sk

Abstract: The number of cyber-attacks on the Internet increases greatly and this trend does not seem to stop any time soon. The spread of malware is fuelled by various factors, with the technology and Internet access becoming relatively affordable globally and forced home-office of various professions, which increases the possible threat exposure time. Threat actors also employ various attack vectors, often utilizing some form of position deception to hide their activity from the authorities. Depending on the attacker's skillset, motivation and available resources, the attack may prevail into successful data breach, theft or data integrity violation. These outcomes may sometimes have tragic consequences. Without access to any kind of private data banks, this work was limited only to publicly available sources alongside with their drawbacks. This paper proposes a tool which is able to accept various sources of data, be it providers of suspicious IP address lists, lists containing IP addresses that are known to be part of any kind of TOR/VPN network, blocklists that contain various data and lastly also geolocation databases as means of gathering intel about IP addresses that are either part of suspicious lists or inside of manual queries. The proposed tool was then tested on publicly available data and the results, originating mainly from generated maps and graphs of various categories, alongside with the actual tool were compared to other cyber-threats origin information services and to other statistics about the situation in the online field.

Keywords: geolocation; threat origin analysis; cyber threats, public sources; public lists; information gathering; suspicious IP addresses

1 Introduction

With the continuing trend of more and more devices being connected to the Internet, the risk of exposure to some kind of cyber threat increases, even though digital skills of Internet users increase, and the number of Cyber security solutions is on the rise. The main factor to consider is that, even if only 1 ‰ of users was subject to successful attack, it would mean that 1 thousand people out of 1 million became victims to some kind of malware, phishing, scam, information leak, digital integrity

of data violation, etc. Reasons, for which threat actors often utilize same or similar attack vectors might come with increasing difficulty that is needed, to overcome security obstacles utilized by modern computers, or simply, because attack vector reusage, with minor personalization can still affect a great number of devices, especially those who still employ inadequate or zero protection.

Geolocation is a concept providing means of adequate geographical place of origin identification. The ability to track a country, in case of a bigger agglomeration also a precise state or a city may help a company selling advertisements, a website that adapts its language to visitor's expected language or to follow certain legislation and law requirements in different states.

However, some people's intentions are not always lawful, and their goal may be to steal, destroy, modify, control or in other way disrupt the device's intended operation, valuable information or even a whole company. Their motivation may be of pure greed, envy, revenge, but it can also be a thing of political and technological warfare. There are certain parallels that can be found even in the work *Art of War* [14], written by Sun-c'. And similarly, as we can observe the win of a trick over brute force in the legend of *Troy* [15], hackers often utilize means of social engineering or combination of various attack vectors, with great impact.

Devices are becoming more affordable and the number of connected devices to the Internet grows each day. This growth was accelerated also due to the fact that during Covid-19 pandemic, many companies moved their employee's daily operation online, greatly increasing the risk of their exposure to certain dangers, which did not go unnoticed by the attackers, further expanding their operations.

Their impact would be highly limited if they could be tracked easily, so they employ various evasive [11] and counter-detection techniques, to secure the highest up-time. In case that their goal is to demand a ransom for an unblocking of victim's computer, which is not always a guarantee as a deciphering part of the malware may not be present intentionally, they have moved the payments into crypto world, where payments appear to be more anonymized.

The aim of this paper is to get an insight into the Cyber security situation, or as it may be, its appearance through information that is publicly available from various sources. Precisely speaking, discovering the most notable countries of origin that appear to be the source of selected attack types, with an attempt to further improve the results with other findings, performing automatic data fusion, resulting in enriched outputs, providing outputs in the selected form, with the help of the designed tool, utilizing proposed algorithm that is visualized in the diagrams. Further goal is to offer an evaluation of generated results, creation and later description of figures and tables that we deem interesting, while also taking one, the publicly available nature of the data and second, the size of country size into account. The expectation and a major challenge of this research is that even if we had all the available data, not just public, it would still render the outputs incomplete, as many (especially ongoing attacks) attack vectors, threat actors and agents are yet to be discovered, if ever.

2 Related Work

There are multiple Cyber Security solutions providers, which make and deploy honeypots, IDS, firewalls, antiviruses, anti-spam, botnet detectors, etc. Gathered data is then used for further analysis and security optimization of their products or adapting solutions for their customers. Their data is not usually publicly available for further use, yet some of these providers publish visualizations with anonymized data showing current/historical situation about ongoing attacks, countries of origin/target with utilized ports, allowing for some result filtration [9].

Among those that we found and deemed most interesting are *Digital Attack Map* [1], which also shows important notice about current large and unusual attacks, and also *Talos* [8], which is a visualization from Cisco. Apart from the general view, it also gives a summarized view on top 10 spam and malware senders, with granularity focused on organizations and countries.

Other works focus mainly on using IP geolocation as means of pure blocklist adaptation or helping law forces in an attempt of investigation to hold certain criminals accountable, when the Internet Service Provider (ISP) can be contacted with the obtained geodata, to get the Network Address Table (NAT) mapping, incident logs to confirm the time accuracy of the occurred events, which can then lead to getting the Media Access Control (MAC) address of the host, their real name, real address [5].

3 Tool Design

Due to the technological limitations such as lack of live data about suspicious activity caught from honeypots, which are generally reserved and kept private by Cyber Security companies protecting certain institutions [7], it is difficult to get a true perspective of the situation, which gets even more complicated when we consider the fact that skilled criminals employ various means of protection, such as: Virtual Private Network (VPN) services [12], hijacked servers, botnets, The Onion Router (TOR) routing [13], cellular data from Subscriber Identity Module (SIM) cards not fixed to their Identity Document (ID) cards, temporarily paid hosting services [2].

We can explore the Internet and search for various sources that do claim to have discovered this information, but since we focus on publicly available (and free) resources, the accuracy of this data is limited. Nonetheless, approach that we propose aims to make use of these, to an extent, unreliable information [4] and provide an insight on the current situation and compare it to other available statistics, then judge the results.

3.1 Focus on Attack Types

It is important to mention that these categories are grouped into suspicious lists represented by their respective instances and tagged accordingly. Further geolocation and information gathering can be performed on the whole list by performing one action. If the list is too big, it is divided into batches that are then geolocated according to the limits of geolocation services presented in the system, which is further elaborated later in this article.

Dridex, QakBot, Emotet

Separate tags in the system. Found blocklist return data about IP addresses connected with attacks utilizing these malware types, almost entirely cryptocurrency ransomware [3], [6], [10].

Botnet

A found list of IP addresses that claims to hold information about IP addresses that were part of botnet. No more information is known.

Spam

A found list of IP addresses that claims to hold information about IP addresses that are suspicious of spam activities targeted on forums. No more information is known.

Mail

A found list of IP addresses that claims to hold information about IP addresses that are suspicious of attack on mail servers. No more information is known.

Resilient

A found list of IP addresses that claims to hold information about IP addresses that are online for at least 5 weeks and with at least 5000 recorded attacks are tied with them. No more information is known.

Brute-force

A found list of IP addresses that claims to hold information about IP addresses that are suspicious of brute force attacks, cracking passwords on websites, etc. No more information is known.

3.2 Geolocation

When it comes to geolocation, it is important to track the number of allowed requests per, e.g.: minute, week, month. In this work a minute interval was chosen globally for all the geolocation databases. Therefore, researcher needs to provide the system with correct information about the valid limits of the added provider. Empirical experiment is also recommended for some thresholds observations. Then also, the way how to request a response is needed. Names of the fields and format

in which the response is received is also required, in order to map them to fields in our database. We have chosen such Application Programming Interface (API)s that cover as much information as possible, as some services are not providing, e.g.: Autonomous system number/name or other provide us with information whether the IP is hosted, uses cellular data, … That will be another aspect of the provided detail alongside with other findings.

3.3 Blocklists

Blocklists generally provide a list of IP addresses that can be used as means of website protection. It is important to update them as, regular owner can regain control after some time, etc. These lists may contain tags as reason of presence in the list or even some other useful fields: country, Autonomous System (AS), etc.

3.4 General Idea of Outputs

The thought behind the output is that the viewer visits a certain page that is of three types. A graph, map, or an individual/fused output.

3.4.1 Individual/Fused Outputs

The term *individual* and *fused* always describes one IP address that underwent geolocation procedure, either by manual request via form, or selecting certain suspicious list (or its part). Visitor is always able to see more information about the sources of this data/findings. Individual means that information is provided only by one geolocation provider and fused means, that data is acquired and joined from all geolocation providers. The data is complemented with other module findings (cover lists, blocklists, suspicious lists).

3.4.2 Graphs

Graphical view, its data is issued for the visitor by his browser fetching specific files, generated by the system every N second (e.g.: 60 seconds, interval that can be increased/shortened) from data that is publicly available and present in the system. Chosen graph categories are *top* and *comparison*. The idea behind TOP is that the graph shows top 5 values in graphical way, where the bar type can be changed dynamically, and top 500 values presented in a table under graphs, in case more detail is required. The idea behind comparison is that we identify certain interesting topics, in which we can specify values up front, and the interesting thing is the difference in metrics that is observed in those values. In the top graph category unique values have to be identified dynamically, therefore, the limit for top 500 values in table is presented.

Top

- *Origin* – countries that appear to be used as threat actor's source most often.

- *Signatures* – biggest occurrence of signatures present in the lists in the system.
- *Ports* – ports found to be used most often in attacks.
- *AS* – autonomous systems out of which the threat actors seem to originate.
- *Tags* – types of suspicious lists and IP addresses used in geolocation.

Comparison

- *Online vs Offline* – number of IP addresses available vs already down [2].
- *Disguise* – number of IP addresses that use (and which) technologies to hide their real position by using another IP address(es) and those that do not.
- *http(s)*– number of threat actors that use SSL and those who do not.
- *IP vs domain* – number of IP addresses which operate under a domain, likely to act as someone more trustworthy, perhaps to act as some other similar domain.

3.4.3 Maps

Maps are visualized on interactive 3D model of Earth, where similar categories are shown with the added visualizations of ransomware, countries of origins, etc. Identified points are placed on the model, where they can be clicked on to get more information about them. This can be later used in documenting the outputs of this academic work and comparing with other publicly available data:

- *Specific origins for Dridex/Qakbot/etc.* – notable often occurring world points.
- *Specific origins for suspicious lists* – most occurring locations of each type.
- *Other categories chosen from Top* – places of origins for some categories.

3.5 Presented Algorithms

Every algorithm presented in this part is abstract. The exact implementation varies in detail and is different in a way that, e.g.: data about VPN/TOR/etc. are being looked for in *cover lists.* Other lists follow respectively in general, depending on the created graph/map files topic of interest. When performing suspicious list geolocation, it is a case when the actions of which contained IP addresses are suspected of are clear. Nonetheless, geolocation has to be performed regularly and exploration for other findings is performed as well during analysis. Another case is when the list input is manual. Then, no such information is available beforehand and exploration for potential match is performed for each list respectively alongside the regular geolocation. An optimization is employed, where no deep analysis is performed until the last batch of IP addresses is processed.

3.5.1 Geolocation Algorithm

Algorithm in Fig. 1 shows how various APIs and their limits are utilized, how lists are split into smaller batches that are then requested, processed and saved.

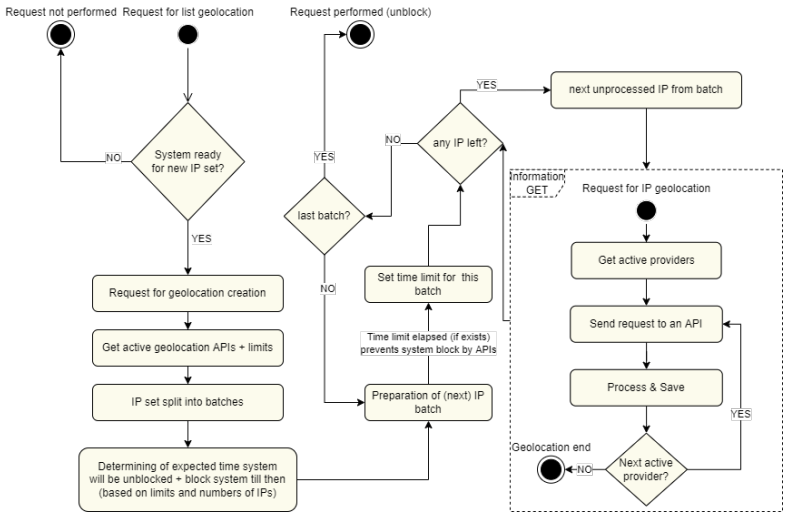


Figure 1
Algorithm showing geolocation requests for APIs

3.5.2 Files for Graphs/Maps Orchestrator Algorithm

This algorithm in Fig. 2 shows that the whole idea is to define event listeners and actions, which generate files for graphs and maps. Then, events are being fired regularly, based on the configuration.

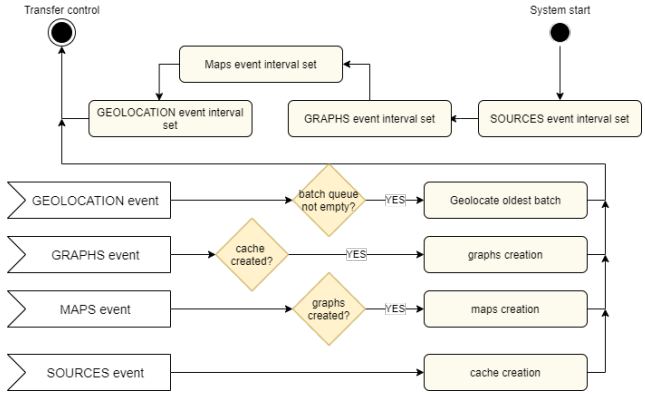


Figure 2
Algorithm that shows regular updates to files for graphs and maps

3.5.3 Findings Locator

This abstract algorithm in Fig. 3 is responsible for matching information from sources/lists present in the system, resulting in returned data, which is then put into file(s) that are utilized by graph/table/map engines.

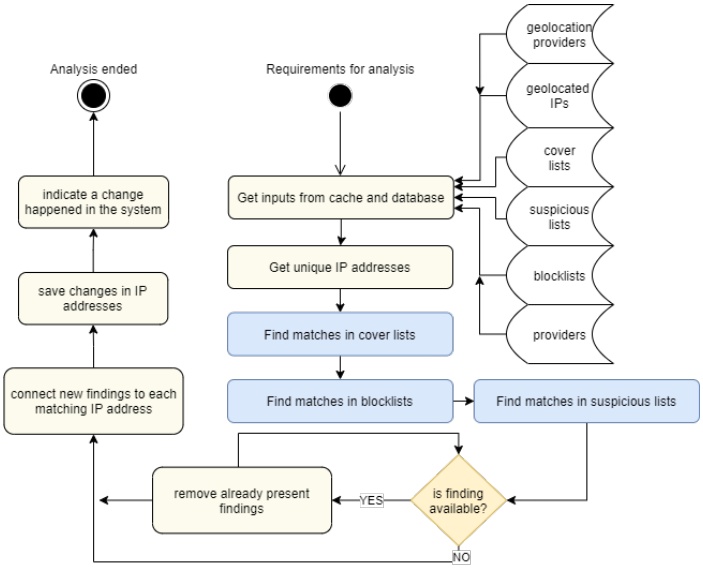


Figure 3
Algorithm showing process of looking for matches in various lists for unique IPs

4 Results

In this part, notable parts of the system outputs are presented, in a way that attempts to provide a meaningful view into the current state of cyber threats in the world, as observed based on acquired and utilized lists, sources and data.

4.1 Used Data

Every source used for analysis, has the nature of being published free of charge. The nature of those sources is, therefore, relatively unreliable. The true potential of the research tool can be, therefore, achieved only by having multiple sources of data that could prevent a great bias. However, as there is no such way on how the trustworthiness of used data can be tested separately, the only possible way is therefore to perform a test and then compare and evaluate gathered information.

4.1.1 Geolocation Services

Chosen attributes of requests are covered by two geolocation services, its APIs contribute to the database. The first service is able to provide 45 IP addresses per minute, the second service has no such limitation. Empirically it was proven that only 3 IP addresses per minute are valid. The system therefore chose 3 IP addresses

per geolocation event. The issue of lengthy lists for analysis is mitigated by implementation of batch pre-insertion control whether the IP address was already geolocated and, except for IPv6 addresses, every IP address is grouped into /24 CIDR subnets, with expectations of observing similar/same geolocation position, thus saving requests:

- IP-API https://ip-api.com/
- IPWHOIS https://ipwhois.io/

4.1.2 Suspicious IP Address Lists

It was not an easy task to choose which publicly available data should be used as a basis for this experiment, as there are some factors that influence these decisions greatly. It needs to be regularly updated, contain sufficient amount of data, while also keep the number of false positives to a minimum. We have chosen to follow a path, where we trust a reputable service [16] that first receives information from Fraud/Abuse specialist, whose servers are often attacked, as a source of IP addresses that are suspicious of performing attacks. We believe that since its data is pre-filtered with reputable whitelists, this data met our requirements.

The system was enriched with lists that are internally marked with tags and all IP addresses in them were subject to geolocation and further analysis. The respective date at which the data was recent is circa 3rd May of 2022 [16]:

- VOIP, SIP, SIP server attacks: https://lists.blocklist.de/lists/sip.txt
- Brute force logins: https://lists.blocklist.de/lists/bruteforcelogin.txt
- Mail, Postfix service attacks: https://lists.blocklist.de/lists/mail.txt
- REG-Bots, IRC-Bots, spam: https://lists.blocklist.de/lists/bots.txt
- Threats responsible for more than 5000 attacks, still online after at least 2 months of activity: https://lists.blocklist.de/lists/strongips.txt
- Attacks on FTP service: https://lists.blocklist.de/lists/ftp.txt

4.1.3 Blocklists

Following sources were added to the system, according to empirical response from these blocklist providers. Respective lists were obtained, out of which interesting attributes such as URL, availability status, IP address, user port, threat reason, were acquired and transferred to the database:

- Feodo Tracker – project of abuse.ch organization that aims to share IP addresses of botnet C&C servers that are responsible for Dridex, Emotet, QakBot, Trickbot, etc. malware family types.
- URLhaus – abuse.ch project that shares malicious URL addresses throughout which a malware of respective family type is delivered.

4.1.4 Cover Lists

There is a solid assumption that skilled attackers utilize one way of position deception as a protective measure in hiding their identity, or another. The identified techniques go as VPN, TOR routing, using Hosting services and cellular connectivity. One of the used geolocation databases provides our system with reasonably sufficient data about the IP addresses nature when it comes to hosting, cellular connectivity, but comes with only joined information whether proxy/VPN/TOR was used without any kind of distinction in between them. Therefore, providers and lists that could enrich our results are of VPN servers and TOR exit nodes IP addresses. The utilized lists in this testing are as follows:

- ProtonVPN – supposedly a list of servers, their IP addresses, of the service ProtonVPN. It is available from GitHub repository, where it continues to be updated in the regular manner and is processed as plaintext https://github.com/X4BNet/lists_vpn/blob/main/ipv4.txt
- NordVPN – supposedly a list of servers, their IP addresses, of the NordVPN service. It is available from GitHub gist file, where it continues to be updated regularly and is processed as plaintext https://gist.github.com/JamoCA/eedaf4f7cce1cb0aeb5c1039af35f0b7
- Tor-IP-Addresses – a list of continuously updated exit nodes of the TOR network available in the GitHub repository https://github.com/SecOps-Institute/Tor-IP-Addresses/blob/master/tor-exit-nodes.lst

4.2 Presented Outputs and Evaluation

The outputs of the system were chosen and put into this part of the work as a way of grouped presentation with the goal of important points summarization that can be further evaluated and compared. In the Fig. 4, the example of interactive 3D map output is presented to portray the way the visualization with clickable Points of Interests and Tabs separating context, containing the view for other attack types, works. Hence, it also serves as a distribution visualization of discovered threat actors. Other map outputs are shortened and put into tables in this work.

Fig. 4 shows that among used data, spam attacks were detected to originate predominantly in the USA followed by Russia and this list goes on with Indonesia, Germany, Ukraine and United Kingdom. The visualization utilizes the WebGL Earth service, where on-map points are clickable and provide more details.

Fig. 5 shows top 5 autonomous systems under which IP addresses suspicious of malicious activity, based on the utilized data, belong. Russian ISPs placed first and fifth, while other places are occupied by USA hosting services.

Description	Country	Count
bots	US - United States	8216
bots	RU - Russia	5134
bots	ID - Indonesia	1024
bots	DE - Germany	530
bots	UA - Ukraine	520
bots	GB - United Kingdom	512

Figure 4
Map visualization that shows distribution of 4th tab, IP addresses acting as bots

Figure 5
Graph showing Top 5 discovered autonomous system origins

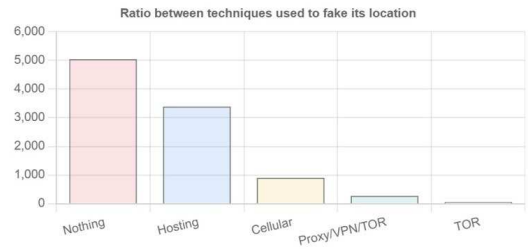


Figure 6
Graph showing Top 5 discovered position deception techniques in use

When it comes to position deception techniques utilization, in the Fig. 6, results from the discovered data can be interpreted in a way that majority of suspicious IP addresses are not using any kind of position camouflage. However, it does not mean, that the threat actors are not using any kind of protection, only that none was detected or they itself were misused for an attack in the form of agent. When it comes to detected deception techniques, the most popular seems to be Hosting, followed by cellular mobile, most likely being just a burner SIM card with pre-paid data plan, and then joined Proxy/VPN/TOR with no way of distinguishing in between them. TOR exit nodes are closing the graph. Specific VPN servers were also linked to some IP addresses, but the quantity was not big enough to overcome TOR exit nodes, for them to be shown in the Fig. 6. The number of IPv4 addresses found to be malicious dominate the graph comparison in the Fig. 7, where threats with IPv6 addresses almost do not exist, while there exists a certain amount of domains that serve as malicious data provider. It is important to mention that this is caused by the small ratio of IPv6 data in suspicious lists.

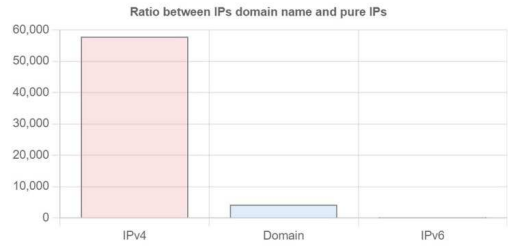


Figure 7

Graph showing the dominance of IPv4 addresses in the included blocklists

Fig. 8 shows unique threat origins (including addresses that use position deception). It can be seen that USA and Russia mostly continue to be seen as the predominant countries used as threat origin, with valid question that resonates:

"Whether this order is a coincidence, or if the situation comes from the fact that some countries have problems with regulation of cyber threats due to their size or possibly due to strategical political motivations."

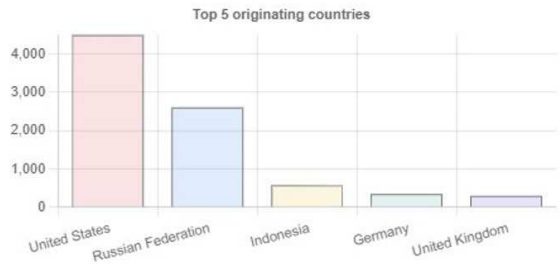


Figure 8

Graph showing Top 5 discovered unique origins of threats

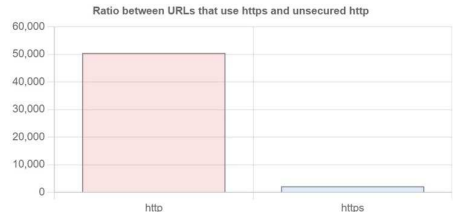


Figure 9

Graph showing the usage dominance of http over more trustworthy https protocol

In Fig. 9 the comparison of *http vs https* usage protocols is shown. It is necessary to mention that for the informed user, this alone might be suspicious. Only 4.12% of threats use https for their malicious activities and attempt to hide the traffic sent.

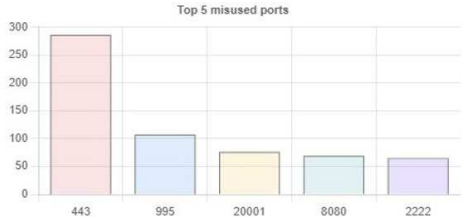


Figure 10

Graph showing Top 5 ports discovered being misused by the attackers the most

From Fig. 10, the prevalence of port 443's misuse is evident. Rest of the ports displayed in the graph continue with a lower, but to themselves similar quantities, trend of which continues even with ports that are not shown in Top 5.

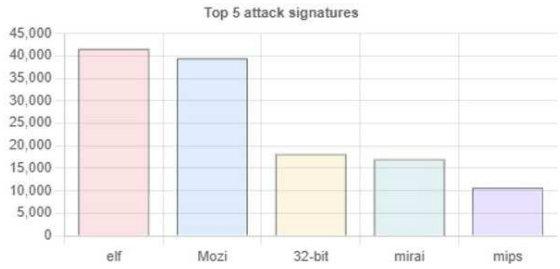


Figure 11

Graph showing Top 5 threats tags present in the system blocklists

Fig. 11 shows the quantity of threats that are present in the system. It may be surprising to see threats aimed at Linux operating systems (elf) in the first place, which contradicts the popular belief of virus free experience. Dominance of botnet malware families (Mozi and mirai) is also present in the graph alongside Linux threats. Less common, but still very potent threats include malware that targets IoT devices specifically (mirai and mips). Among others, there is a great number of threats consisting of shell codes built for 32-bit operating systems.

In Fig. 12, comparison of current threat activity status is shown. The fact that a threat is offline, does not necessarily mean that the URL or IP address is no longer reachable, it may mean that the rightful owner has regained access, in other words, an address no longer serves as threat agent. The majority of threats present in the system are offline. Therefore, if any kind of publicly available (e.g., via REST API) blocklist would be provided by this tool, only IP addresses of threats that are online should be considered 88.38% of threats present are offline.

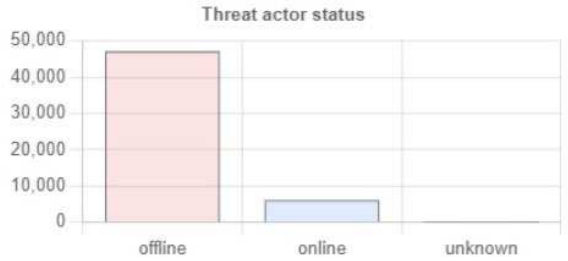


Figure 12

Graph showing that the majority of threats in the system are already offline

Another continuation of figures, showing data from maps could be presented, but it is decided, that presentation in the form of tables may be of more comparison value. The following Table 1 shows the trend of top country order occurrence, listing the possible means of position deception, linked in the data fusion. Based on detected usage, USA Hosting services were found to be the most attractive to threat actors, possibly due to their availability and simplicity of use. Popularity of this technique was also found in Singapore, Germany, China and India. Cellular services were most often misused in USA and Russia, while TOR exit nodes, seemingly random, were found to be used most often in Luxembourg. For some data, the system was not able to clearly identify which technique was used, placing it into joined category Proxy/VPN/TOR, which also showed that the most prevalent origin is in USA and China. One of the major factors contributing to the repeating appearance of certain countries in tables is that not only many providers with massive networks operate there, are as well many companies offering commercial solutions to customers from around the world.

Table 1

Table showing top 10 of grouped position deception techniques identified to be used by threat agents, with the exception of Nothing detected, which was omitted from this graph

Order	Deception technique	Country
1.	Hosting	US
2.	Cellular	US
3.	Cellular	RU
4.	Hosting	SG
5.	Hosting	DE

6.	Proxy/VPN/TOR	US
7.	Proxy/VPN/TOR	CN
8.	TOR	LU
9.	Hosting	CN
10.	Hosting	IN

Table 2 contains information regarding order of botnet malware family type occurrence in a certain country. Majority of the findings belong to QakBot family type with countries like USA, United Arab Emirates, Mexico, India, Brazil and Pakistan being the supposed C&C server botnet origin. Emotet malware family is represented by USA and France while Dridex by USA and TrickBot by Columbia. Further listing of each category would be possible, but this is the chosen view which shows these threats ordered in quantifiable manner as seen in experiment.

Table 2
Table showing top 10 botnet malware families alongside with countries of origin

Order	Botnet family	Country
1.	QakBot	US
2.	QakBot	AE
3.	QakBot	MX
4.	Emotet	US
5.	QakBot	IN
6.	QakBot	BR
7.	Emotet	FR
8.	QakBot	PK
9.	Dridex	US
10.	TrickBot	CO

Table 3
Table showing top 10 threats marked by suspicious lists as ordered by their quantity alongside the supposed country of origin

Order	Threat type	Country
1.	Bots	US
2.	Mail	US
3.	Bots	RU
4.	Mail	RU
5.	Bots, Mail	ID
6.	Bots	DE
7.	Bots	UA
8.	Bots	GB
9.	Mail	DE
10.	Strong	CN

Table 3 shows that majority of threats are aimed towards forums performing malicious spam operations, while being followed by attacks aimed at Mail servers that are both sharing the same origin similarities. USA, Russia, Indonesia, Germany, Great Britain and Ukraine were found to be origins of these threats, with resilient IP addresses of China closing the top 10 table.

Conclusions and Future Research Directions

With the use of the designed and implemented software tool, experimental research was performed, with outcomes showing comparable level of accuracy to published data, confirming the leading positions of USA, Russia and China as countries with the most prevalent probability of being a threat actors' true origin or at least origin of the misused threat agents. Despite this, it cannot be confidently expressed whether this is a true state of the situation, as many position evasion techniques are utilized and even though they were, in lots of cases, identified, the true geographical location could not be determined. Other types of information that the tool provided are believed to be mostly accurate, the only issue being that the input data does not contain information about all types of attacks, but for the sake of keeping false positives to a minimum this set was chosen. The provided insight was created with the help of data fusion, from data that was all publicly available, parsed and filtered.

An important notice is that the generated results presented in this work serve mainly as a snapshot of the situation in May 2022. Ongoing monitoring would probably show changing trends in the most prevalent countries of threats' origins, but we believe that countries like USA, Russia and China would remain at the top. New types of threats could emerge, and they would be observed, as information from blocklists is being updated regularly. When it comes to commercial companies providing free informational visualizations online, the data is as well ever changing. But with limited amounts of mostly anonymized detail and often shown spikes of malware senders in certain countries as Brazil, Vietnam, Hungary, etc., it is not easy to do a thorough comparison. The main advantage they have, is that the data is under their control, giving them the ability to perform more advanced research internally [1], [3], [6], [8].

The strength of this tool is also its main issue. Data is freely available on the internet and even though the provider is generally considered trustworthy [16], it is not a guarantee for other providers, where it can be mostly just assumed. If someone would purposefully change the IP addresses in the public lists, or created others that this system would find and utilize, they would be shown as a potential threat, resulting in a false positive, which if combined with indication of online status could result in them being part of a future blocklist. Would the data in the lists was instead supplied with a trustworthy honeypot, suspicious IP addresses would be only analysed for their attributes and potential cover methods, greatly increasing the trust in results.

As one of the most important concluding remarks, that should be mentioned is, that greater accuracy to the true situation can be achieved only if various lists of

suspicious IP addresses are present in the system, as one type of attack might be more popular in one country than in the other one. It would display a biased information in graphs and this needs to be eliminated to a minimum with a variety, which gets complicated with free sources. Further blocklists that focus on manifold threats and lists focusing on single threats would contribute greatly to the overall results. Accuracy of the outputs is also increased, the more information about servers behaving as position deceptors, again, is present in the system. This would be best served through an integration with some other non-public solution.

The implemented data processor could be suitable for evaluation of future results in an ongoing manner, due to the nature of information liquidity in blocklists, etc. which can be regularly updated automatically. This system may enhance the security of websites that want to increase their security and use a joined blocklist of IP addresses that were found to be suspicious in any way, that could be present to an endpoint via API.

It is necessary to take the outcomes of the implemented tool with a pinch of salt, but its ability to provide educational level of information, in a transparent manner of visualizations in graphical way was achieved and its outcomes were utilized.

The aim of this paper, to provide an insight into the Cyber security situation or as it may appear through the publicly available data is deemed as fulfilled and the tool shall stay in use, yet possibly utilizing premium geolocation services that offer more accurate and up-to-date data, with greater number of requests, possibly allowing much greater throughput. As mentioned earlier, setting up this tool to receive data from some honeypots is also a plan. Few remarks came during the research that deserve to be mentioned here, although answers to those questions are yet to be discovered:

- Is it possible to distinguish between attackers of various experience, skills, motivation and utilized resources on attacks? Did they unleash their maximum potential from qualitative and quantitative perspective?
- Is the origin of the IP address the system shows, when no position deception is discovered, the actual origin and in no way covered with different position deception technique? Or are there any other techniques left yet to be discovered?
- Is the IP address figuring as cyber-threat origin a true threat actor or just a bot in a botnet, when no information about presence in the botnet is discovered?
- Can it be assured that the IP address of the threat agent is still under threat actor’s control?
- Was the attack successful? How much of the attacker’s intentions were achieved? What was the target of the attack? What were the consequences and the harm caused?

The actual usage of publicly available data comes with certain risks of false negatives or false positives – as it might take some time until update in acquired blocklist occurs. Furthermore, this tool does not aim to render paid services obsolete, but to provide an alternative to them. Future elaboration of performance, data accuracy and actual potency to request more IP addresses per minute, may show that access to paid geolocation services is required.

Acknowledgement

This work is a partial result of the Operational Programme Integrated Infrastructure for the projects: Research in the SANET network and possibilities of its further use and development (ITMS code: 313011W988), CEVIS Support of excellent labs research activities STU Bratislava (ITMS code: 313021BXZ1) and ACCORD Advancing University Capacity and Competence in Research, Development and Innovation (ITMS code: 313021X329), co-funded by the European Regional Development Fund (ERDF). This research was also supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic, Incentives for Research and Development, Grant No.: 2018/14427:1-26C0.

References

[1] Digital Attack Map, 2020 [online]. 2020-01-01 [visited on 2021-05-16]. Available online: https://www.digitalattackmap.com/

[2] Newman. L., 2017. How an accidental 'kill switch' slowed Friday's massive ransomware at-tack [online]. 2017-05-13 [visited on 2021-03-02]. Available online: https://www.wired.com/2017/05/accidental-kill-switch-slowed-fridays-massive-ransomware-attack/

[3] Kuzmenko, K., 2021. QakBot technical analysis [online]. 2021-09-02 [visited on 2021-12-04]. Available online: https://securelist.com/qakbot-technical- analysis/103931/

[4] Komosny, D., Vozňák, M., Rehman, S., 2017. Location Accuracy of Comercial IP Address Geolocation Databases. Information Technology and Control. Vol. 46. Available from doi: 10.5755/j01.itc.46.3.14451

[5] Hillmann, P., Stiemert, L., Rodosek, G., 2020. Dragoon: Advanced Modelling of IP Geolo-cation use of Latency Measurements. Available from arXiv: 2006.16895 [cs.NI]

[6] Grammatikakis, K., Koufos, I., Kolokotronis, N., 2021. Understanding and Mitigating Banking Trojans: From Zeus to Emotet. In: 2021 IEEE International Conference on Cyber Security and Resilience (CSR), Available from doi: 10.1109/CSR51186.2021.9527960

[7] Wang, H., Gu, J., Wang, S., 2017. An effective intrusion detection framework based on SVM with feature augmentation. In Knowledge-Based Systems. ISSN 0950-7051, 2017, Vol. 136

[8] Talos, 2020 [online]. 2020-01-01 [visited on 2021-05-16]. Available online: https://talosintelligence.com/fullpage_maps/pulse

[9] Zima, S., 2016. Analysis of geolocation databases. MA thesis. FIT VUT Brno

[10] Rudman, L. Irwin, B., 2016. Dridex: Analysis of the traffic and automatic generation of IOCs. In: 2016 Information Security for South Africa (ISSA), Available from doi: 10.1109/ISSA.2016.7802932

[11] Bukac, V., 2010. IDS System Evasion Techniques. MA thesis, Masaryk University, Czech Republic

[12] Berger, T., 2006. Analysis of current VPN technologies. In: First International Conference on Availability, Reliability and Security (ARES'06), Available from doi: 10.1109/ ARES.2006.30

[13] Bee S., 2020. How Does Tor Really Work? [Online]. 2020-09-19 [visited on 2021-05-13] Available online: https://skerritt.blog/how-does-tor-really-work/

[14] Carlsson, A., Gustavsson, R., 2017. The art of war in the cyber world, 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), 2017, doi: 10.1109/INFOCOMMST.2017.8246345

[15] Maher, M., 2003, Fall of Troy VII: New Archaeological Interpretations and Considerations, TOTEM: The UWO Journal of Anthropology, TOTEM Vol. 11 2003

[16] Blocklist.de, 2022 [online]. Export all blocked Ips, 2022-05-05 [visited on 2022-05-05]. Available online: http://www.blocklist.de/en/export.html

Top Executives on Competitiveness: Survey of Digitalization and Internationalization in Hungarian Companies

Xénia Szanyi-Gyenes, Anikó Almási

Corvinus University of Budapest, Fővám tér 8, 1093 Budapest, Hungary;
xenia.gyenes@uni-corvinus.hu; aniko.almasi@uni-corvinus.hu

Abstract: The "Future of Organisation and Leadership" research project, explored what leaders of companies, operating in Hungary, think about their future opportunities and challenges, and how their current competitiveness, can best be supported. The answers to our questionnaire survey, showed that the future companies will be highly digitalized. Therefore, we investigated how companies in Hungary approach the issue of digitalization and how it contributes to their competitiveness. Our results show that a number of factors are related to the future competitiveness and digitization of companies. Managers determine the functional and company-level strategic areas that a company engages in digitalization, and this demonstrates a clear pattern with other competitiveness factors. Examples include internationalization, strategic maturity or managerial competencies and knowledge boundaries. The results expose the barriers and the catalysts for the future success of companies in Hungary.

Keywords: competitiveness; Hungarian companies; internationalization; digitalization

1 Introduction

In the research presented in this study, we assessed the current state of Hungarian companies and their expectations for the future. The research team of the Institute of Strategy and Management of Corvinus University of Budapest has been investigating the present and future of the operation and management of companies in the framework of the "Future of Organisation and Leadership" research [1]. Company leaders answered questions such as: What do they think about the leader of the future, their organization, how do they think companies can adapt to changing expectations and environmental challenges? What practices are disappearing, what new knowledge will leaders need to manage the organization of the future and make strategic decisions? How the Hungary-based corporate sector relates to future expectations, what development trends it perceives, what it is doing to "keep up with the times" and maintain its competitiveness. To this end, we conducted a broad

and detailed survey of the managers of Hungary-based companies, focusing on several areas. A sample size of more than 300, provides a good analytical opportunity, from a research perspective. The aim of this article is to better understand the competitiveness of Hungarian companies in the context of digitalization and internationalization. In the present analysis, our aim was to explore the areas in which digitalization is present in organizations, and what impact it has on their operations, competitiveness and future prospects. We used control variables to identify the strengths and weaknesses of companies.

Based on the initial analyses, we discovered some interesting results, in line with the trends. It was already known that environmental factors play an important role in the formulation of corporate strategy and in management decisions. Another important aspect is the stage of development of the company, because it determines the problems and challenges. The perception of managers plays an important role in decisions, the attitude of the manager has an impact on the competitiveness of the company, the perceived business opportunities, the willingness to take risks and to innovate. Having one or more managers in a company also seems to be essential for digitalization, as shared decision-making involves a higher level of data analysis. Similarly, internationalization was a control variable: companies with an international presence tend to have a higher level of strategy and less of an ad hoc decision-making culture, which also favors long-term, data-driven decision-making and corporate digitalization.

The first stage of the evaluation of the results focused on variance, outliers, different patterns and their possible background. Examining the factors that fit the theoretical models yielded standard results without variance, so we started to look for areas that could still capture, characterize and explain the different ways in which companies operate and compete. By comparing the individual factors in a non-model-specific way, we expected to find homogeneous groups of firms that differed significantly from each other. The limitations inherent in exploratory analysis must of course be taken into account: the results presented here await even larger sample testing and more detailed statistical analysis.

2 Literature Review

2.1 The Innovation Aspect of Competitiveness

In the field of corporate competitiveness, there are many parallel approaches (industry, regional, product and product group, business line, etc.). The diversity of definitions has sometimes caused methodological problems [2], but has helped the rapid development of different theoretical concepts. In analyzing the results of the present research, a classical definition of competitiveness is taken as a starting

point: a firm is competitive if it "can consistently offer consumers products and services that they are more willing to pay for than those of competitors on terms that ensure a profit for the firm" [3]. In the present study, we examine the competitiveness of Hungarian companies, focusing on the challenges of digitalization on the one hand, and on the other hand, we analyze the future competitiveness opportunities of companies from the perspective of internationalization.

The role of strategy in the competitiveness of a company has come to the fore, in addition to the soft-hard and resource-capability factor divisions, because it is essentially the company's leader who determines the strategic goals, future plans and the alternatives to achieve them. The influence of the manager in the SME sector is more direct [4] than in a large company, for example, the assessment, perception and development of organisational capabilities is essentially up to him/her [5] [6]. The manager has the power to decide between different options, to determine the way of reacting in case of change and crisis, to allocate the necessary resources. This is the reason why we looked at the role and impact of the leader, mainly from the point of view of strategic decisions. Therefore, different revenue stages were analysed separately to explore the differences between them in terms of strategy, leadership and decision-making.

However, resources are limited for most SMEs, so the efficiency with which they are used by the decision-maker, and through them by the company, is not the same. Digitalisation can be one of the tools to help a company operate more efficiently and effectively. However, digitalisation is not an innovation in itself, it only becomes an innovation if it can be truly classified as an innovation. One of the most widely used innovation typologies is the classification of product, process, marketing and organisational innovation [7] [8]. Digitalisation can be seen in the enhancement of the user value of a product or service, in the product positioning, promotion and positioning strategy related to marketing innovation. Customer database analysis can increase corporate profitability in many innovation areas, as digitisation can provide decision makers with more accurate data and faster access to reports. Organisational innovation is defined primarily as new ways of interacting within and between organisations. Digitalisation can broaden the channels of communication, create new ways of interacting and sharing knowledge, and provide new structures for accessing information by organisational members. Process innovation also focuses on an organisational factor: increasing the efficiency of the internal organisation. Process innovation is often technological and digitalisation can also play a key role in the introduction of new channels, automation of warehousing, tracking, planning and control, and distribution management [9]. Several EU SME competitiveness studies have shown that the digitalisation of firms has become a critical factor in their competitiveness and is expected to further increase its role in firm performance in the near future [10] [11]. According to the WCC/IMD global research [12] and there is the wider environment, the level of digitalisation of society, so digitalisation has impact on all areas of business

operation and management. Therefore, in this study we analyse not especially the innovation, but the digitalisation of firms, focusing on the manager, with control variables such as internationalisation, turnover or the period of the company's establishment.

A recurring pillar of Hungarian competitiveness research is the ability to innovate, to properly assess, develop and exploit organisational knowledge, and to lead in these areas. These approaches also focus on different areas: efficiency [13], creativity [14] [15], the role of IT and controlling and data-driven strategy making [16] [17], and leadership [18] [19]. Innovation or internationalisation appears as a separate factor among the ten factors of the Small Firm Competitiveness Index [20] [21]. The innovativeness of companies has been examined in product, process, marketing and organisational dimensions, but other studies have also concluded that the innovation capacity of the Hungarian SME sector is extremely low [22] [23]. This is also true in relation to other competitiveness factors and in international comparisons [24]. It is regularly argued that the innovation barrier, of a firm, is a leading factor and for the firms that excel in internationalisation and networking, innovation and competitiveness is generally higher. [25] [26]

As shown by the pandemic, digitalisation was a key factor in companies' rapid and effective response to environmental challenges. These rapid responses required a fast decision-making mechanism and digital skill-set in the company. At the center of corporate digitalisation and decision-making, especially at SMEs, is primarily the manager. But business owners and managers usually see IT development as a costly and difficult-to-recover investment. The companies are more inclined to digitise when the business environment changes quickly and unexpectedly [14]. Researches have prooved company performance and profits increase significantly when development is not done in isolation, but in an integrated way, covering the entire company. It all depends on the IT attitude and the decision of the leader, the (re)allocation of resources. According to a survey of Spanish SMEs a more costly, larger-scale but less integrated TQM system implementations were associated with lower levels of profitability [27]. Other research has found a relationship between company growth and the level of integration of ERP implementation. Analysing 352 Danish SMEs, the researchers concluded that the growth rate of young companies early in their life cycle is strongly dependent on the complexity of the ERP system and the level of integration [28]. SMEs often have less standardised processes and low levels of documentation, so IT developments help to improve the objectivity of corporate decision-making. With digitalisation, start-ups can support higher levels of knowledge sharing, manage more and more efficient information channels and operate in a flexible, data-driven way without having to go to the manager before every decision. Management involvement, engagement and support are therefore inevitable in early digitalisation processes [29] [30].

Innovation is the key driver of corporate competitiveness theories. Research has shown that many other factors, such as internationalization, decision-making mechanisms and strategy, can also be emphasized through the leader in corporate

competitiveness. Therefore, in this paper we focus on these factors in presenting the research findings.

2.2 Internationalization Aspect of Competitiveness

The survey did not define the concept of internationalisation, leaving it to respondents to interpret their company's presence in international markets. "Is your company present on international markets?", followed by four response options: "we are present on international markets; our long-term objectives to enter; medium-term objective to enter; and we do not plan to enter international markets". The concept of internationalisation is sufficiently broad and diversified to encompass virtually all the activities by which a company is involved in international processes. Furthermore, we have examined the issue of internationalisation by ownership background, which companies have a majority international ownership structure.

The internationalisation of companies, is a widely researched area in economics [31-34] [36] [43]. Thanks to globalisation, companies are increasingly more easily involved in international processes. Companies entering international markets are more competitive and open to innovation. Small businesses are also increasingly asserting themselves in the international market, and are rightly the focus of research attention: the study of managerial attitudes in Hungarian SMEs [37]; learning and development in small businesses [38]; the study of factors influencing the survival and growth of Hungarian start-ups [39]. The domestic market offers limited opportunities for growth. Innovation and dynamic growth objectives require international market access.

Theories initially took a macro perspective, and the focus on firm-level factors came to the fore only later [27] [40]. Initially, corporate research focused on large firms, with attention to smaller firms starting in the 1990s, when some researchers identified a firm-size-specific growth trend [41]. The classic theory of corporate internationalisation is described by the Uppsala model, where internationalisation is a step-by-step process of market engagement [31] [42-44]. The Born Global concept is that companies see the whole world as a potential market from the outset [45] and think in international terms from the outset in all their corporate activities. They internationalise quickly and generally successfully [46] and focus on developing the skills needed to achieve success internationally [47], becoming more competitive overall. Business competitiveness research looks at how companies can gain a sustainable competitive advantage and what factors influence their success and competitiveness. The pillars of competitiveness of (small) firms are internationalisation, technology, innovation capability [22], market orientation [48], opening to new markets, export orientation [49] and integration of international knowledge as competitive factors in market competition. There is a strong link between internationalisation and innovation: according to a Chinese study [50],

innovation has a negative impact on the survival rate of exporting firms when the firm is solely domestically owned and has poor profitability. For firms with international ownership and a stable profit margin, the effect of innovation was markedly positive. Other research has also demonstrated the positive impact of innovation on financial performance and highlighted the influence of internationalisation [51].

More competitive companies are more successful than their peers. Ónodi-Répáczki [30] investigated how leadership skills influence the success and efficiency of a company, looking at innovation. The competitiveness of a company is shown by a successful innovation-growth strategy, where the researchers defined success as bringing radical innovation to market [52]. Their research was also conducted among domestic CEOs and found that companies that do not innovate, but only stagnate, already show signs of crisis. Successfully bringing an innovation to market depends on sound technical knowledge, process improvement, the right networking capital and creativity and innovative ideas. Looking at the competitiveness of firms from a different perspective, we find that market orientation has the strongest impact on the competitiveness of medium and large firms, especially when the firm is present in several markets [38], with competitive orientation having the strongest impact on competitiveness among firms in Central and Eastern Europe and Western Europe. In the SME sector in Hungary, even in the case of larger turnover companies, the managerial attitude is pronounced, and the lack of management skills, which have the greatest influence on strategy, operations and process design, is the biggest obstacle to growth and development [53].

3 Methodology

The exploratory nature of the research goal determined the need to develop a complex research design that could respond to changes and unexpected results. Therefore, we based the research stages (research objective, conceptual framework, research question, methodology and validity) on Maxwell's model [54], as it allows for interaction between the elements and has a non-linear structure. The target group of the quantitative research was Hungary-based companies with 50 employees or more and a turnover of at least HUF 100 million. The sampling method was stratified random sampling with defined quotas: the number of employees and the county in which the firms were located. A random method was used to select the firms to be contacted by quota cell, following the distribution of the surveyed firms by turnover and activity. The survey was carried out in 2022, and 305 respondents were in the final database. To make the sample more closely match the quotas, statistical weighting was used. Of the sample, 53.5% are medium-sized enterprises (50-99 FTE), 29.7% are medium-large enterprises (100-249 FTE) and 16.8% are

large enterprises (250+ FTE), which is in line with the proportions found in the overall population.

The respondents were senior managers of the companies: 105 top executives (CEO, Managing Director, Chief Executive Officer), 65 executives subordinate to the top executive (CXO level) and 115 executives for strategic decisions participated. On average, the respondents had 25 years of work experience. The sample reflects the activity distribution of the whole population, with manufacturing accounting for a third of the domestic business sector. 34.9% of the enterprises are located in Budapest and 13.5% in Pest County, i.e. almost half of the sample is located in the Central Hungary region. 74.6% of the companies are majority Hungarian privately owned, 7.5% are majority Hungarian state or municipality owned and 17.9% are at least 50% foreign owned.

4 Results of Empirical Research

4.1 Factors Affecting the Competitiveness of the Hungarian Business Sector

The research focused on a deeper understanding of the competitiveness of Hungarian companies. Based on previous research, we focus our findings on the competitiveness-enhancing impact of digitalisation and evaluate our results in three aspects:

(1) Company turnover

(2) Internationalization

(3) Leadership and decision-making

4.1.1 Results by Category of Company Turnover

In the first dimension, we examined the factors based on the turnover of the companies. We had not previously assumed better results for company size. Almost half of the companies have a turnover of over HUF 2.5 billion. 63% of these companies are classified as medium and medium-large sized enterprises and 37% as large enterprises. In the second category, which is interesting in many ways, there are 76% medium-large sized enterprises and 24% large enterprises.

Previous research [55] [56] suggests that business success is not always conducive to innovation, and that the openness of managers to innovation [30] shows a significant relationship with successful business innovation. Sometimes, firm owners are not necessarily motivated to make a risky, resource-intensive innovation

investment by stable business results, and they focus instead on secure operations that guarantee profits.

Table 1
Turnover categories

Turnover	What was the turnover of your company in 2021?	N=305	%
1	Between HUF 100 and 300 million	18	5.9%
2	Between HUF 301 and 700 million	45	14.8%
3	Between HUF 701 million and HUF 1.499 billion	51	16.7%
4	Between HUF 1.5 billion and HUF 2.5 billion	47	15.4%
5	Over HUF 2.5 billion	144	47.2%

Five turnover categories were identified in the questionnaire survey. The results were analysed along each category. The pattern is clear that companies in the second turnover category often operate differently from other companies.

Table 2
Turnover categories and strategy

Turnover category	STRAT-1	STRAT-2	STRAT-3	STRAT-4
1	16.7%	44.4%	33.3%	5.6%
2	31.1%	46.7%	15.6%	6.7%
3	15.7%	37.3%	41.2%	5.9%
4	19.1%	36.2%	40.4%	4.3%
5	32.6%	29.9%	32.6%	4.9%

STRAT-1: We have a business plan broken down from the strategy to the individual level and quantified and we regularly review the plan-fact variances. STRAT-2: Most colleagues know the long-term goals of the business. STRAT-3: Only a narrow group of people (e.g. middle managers, managers, or family members in the case of a family business) know the long-term objectives of the business. STRAT-4: The manager is aware of the long-term objectives of the company, but does not share them with the majority of employees.

The results show that technological changes, including the challenges of digitalisation, are more challenging for larger companies and have the least impact on companies in the second revenue category. Their differentiation can be explained by the fact that they are at different developmental stages, which presents them with very different types of challenges, but they do not yet have the organisational characteristics of classic larger companies (established management, organisational culture). Once past the initial growth stage, this second revenue category group is the least interested in exploring and experimenting with new opportunities. It is assumed that after the higher risk period of the start-up period, this is not the focus of strategic thinking. This group of companies typically emphasised meeting local market needs when the issue of internationalisation arose. Accordingly, they may be more focused on their existing practices and markets and less on new opportunities.

Companies in the second turnover category are extremely low on the following improvements: most of them do not plan to innovate in technology, digitize their existing processes, improve management skills, or set more detailed organizational goals at individual level in the next 1-3 years. Basically, they do not feel that their revenue growth depends on the digitalization of their decision processes, and they do not see that their decisions would be more accurate with higher levels of data analysis or at least that the time needed for the decision process would be reduced. The most active in these areas are companies in the lowest and highest turnover categories, which make turnover growth dependent on digitalization and skills development.

The results show that the second revenue category, together with domestic market orientation, represents a kind of resting state: the manager can feel successful with his previous strategy and the operation of the company, as it allowed the company to grow and reach the higher revenue level. There is a lack of motivation to change, to take risks and to be satisfied with the current situation. For companies reaching the additional turnover category, it is clear that growth involves continuous transformation, improvement and innovation. They focus more on internal improvements, digitalization, organizational inhibitors such as knowledge gaps or resistance to change. They are less likely to sit back and enjoy their successes, as they are at a different developmental stage and have sufficient data and previous experience, that growth is not an achieved resting state but a continuous task.

At different stages of the decision-making process, the companies in the research rely on data to different degrees. In smaller companies, the use of numerical analysis is typically used for the initial assessment of the situation, and the use of IT support for subsequent steps is not significant. The larger the turnover, the more important the use of data, in the overall decision-making process.

4.1.2 Results by Internationalization versus Domestic Market Focus

As we have seen in the literature review, internationalization is a factor in corporate success and competitiveness. For this reason, the survey asked whether the company in question is present on the international market, has long- or medium-term objectives or does not plan to enter the international market. The question of internationalization was used as a decomposition factor along several questions in order to identify common patterns between companies present on the international market and those focusing only on the domestic market.

Table 3
Internationalization

Is your company present in international markets?	**N=305**	**%**
We are present on international markets	157	51%
Our long-term objective is to enter the international market	45	15%
It is part of our medium-term objectives	22	7%
We do not plan to enter international markets	81	27%

Of the 305 companies in the sample, 51% of companies, are currently present in international markets, while nearly a quarter of the companies have no plans to enter international markets. Of the companies that do not plan to enter international markets, 21% cited a lack of resources and 12% a lack of expertise or experience. Among the free-word reasons given by companies not planning to enter the international market were: "*We have no such ambition; not compatible with the activity; local market is enough; serving local needs; serving Hungarian population, aiming to cover the domestic market; our service cannot do it*". The reasons were clustered along three categories:

(1) Domestic market aspirations, here by turnover category, smaller companies justified their exclusive domestic aspirations as local, while larger companies explained their exclusive domestic aspirations as Hungarian market coverage

(2) Nature of the activity or service does not allow it

(3) Lack of aspiration. The results also suggest that companies operating only in the domestic market do not continuously improve their strategy and are likely to engage in lower competitive intensity.

The aim of evaluating the results by internationalization was to find relevant differences, patterns and correlations between companies present in the international market and companies that (definitely) do not plan to internationalize. When looking at the dimensions of strategy and internationalization, we can see that there is a marked difference in strategic thinking between companies that are currently present in the international market and those that do not plan to enter the international market. Those that are present in the international market seek to maintain a competitive advantage through continuous new strategic initiatives, with more individual-level planning of strategy, compared to companies that are only active in the domestic market, where it is more typical that the company's management is aware of the company's strategy. Accordingly, information sharing is also asymmetric, with information in the hands of the CEO in the case of companies operating in the domestic market, and information flowing downwards in the case of companies operating in the international market.

Table 4

Innovation attitude

How much do you agree with the following statements about your organization? Our organization…	**N=157**	**N=148**
constantly seeks to achieve sustainable competitive advantages over the competitors through new strategic initiatives.	52%	38%
is excellent in efficiently executing and further improving business as usual.	50%	32%
is excellent in exploring and experimenting with completely new possibilities.	48%	30%

is able to adapt effectively even to significant and unexpected environmental challenges	40%	42%
gives priority to digital technologies, i.e., allocates resources to their application and development	39%	28%

The answers were scored from 1 to 5. Here the proportions who gave the maximum score of 5 are highlighted. N=157 who are present in the international market and N=148 who are not present in the international market.

Companies that are also present in the international market perceive the impact of technology, including digitalization, as stronger and pay more attention to the use of digital technologies, compared to domestic companies that perceive the impact of technological change as less of a challenge. The lack of specific expertise as an inhibiting factor is clearly linked to internationalization. Firms that are also present on the international market are likely to have the necessary specialized expertise, such as sales knowledge, and thus to perceive expertise as less relevant as a barrier. Conversely, companies operating in the domestic market may have a real and existing knowledge gap that hinders them from competing in the international market. Firms in the international market perform better in improving work process efficiency, identifying new practices and allocating resources accordingly.

In the case of internationalization, in addition to market orientation, we also examined the ownership structure of domestic firms and found weak but significant relationships in the following areas. A significantly higher proportion of Hungarian firms with majority foreign ownership (more than 50%) perceive that technological change and digitalization challenges have a major impact on their organization compared to Hungarian-owned firms. In their perception of their strengths, they described themselves as outstanding in digitizing their processes and improving their efficiency, and in exploring entirely new opportunities. Compared to their peers, they consider organizational innovation and the use of agile methodologies to be of paramount importance and prioritize digital projects. They also adapt their organizational structures, information flows and decision-making processes to this end: they are outstanding in their bottom-up approach to technological innovation, idea generation and implementation of innovative ideas. They place the greatest emphasis on making data from corporate information systems available to an ever-wider range of users. In contrast, among Hungarian or predominantly Hungarian-owned companies, it is typical that numerical analyses are only available to a narrow group of people, typically managers. They feel that the abundance of data slows down rather than supports decision-making processes, and that they often have to wait for more and more data to make decisions. Meanwhile, companies with an international ownership background are taking advantage of data analytics and can reduce the time needed for decision-making processes by analyzing data in detail.

The coexistence of internationalization with innovation was also striking in general. Companies already operating in the international market are much more likely to adopt the latest available technologies. By contrast, companies producing for the domestic market are less likely to be involved in the monitorization and introduction

of new technologies, and those interested in innovative solutions tend to favor technologies that are already proven, i.e., risk-free and tested. Controlling for turnover, it appears that companies in the highest turnover category are more inclined to adopt new technologies immediately, but buy off-the-shelf solutions and are less likely to experiment internally. The lower revenue categories have a higher rate of experimentation, but the sample is still dominated by firms that finance only proven and tested technologies, with a high rate of 80%. Larger firms are quicker to purchase mature technologies due to their capitalization, while smaller firms are more likely to correct their financial shortcomings through experimentation.

4.1.3 Results by Leadership and Decision-Making Process

We examined the information acquisition and flow capabilities of the firm, with a focus on the development of information systems and data quality. It is important to know who has access to data in the organization and whether the digitization of the decision-making process supports the delegation of decisions to a lower level. The role of data in the decision-making process and other characteristics of managerial decisions (intuition, role of emotions, etc.) were also important. Information and decisions were also a key issue in the management culture question group: avoidance, procrastination, goal orientation and communication were also examined. Based on these factors the database was analyzed with the focus of the leader. We looked at how digitization and innovation in systems that support the flow and sharing of information affect decision-making.

The overwhelming majority of respondents to the survey considered that the organization of the future is one that is technologically advanced, constantly evolving and innovative. This is related to the theoretical direction, which envisages flatter and networked organizations, as IT developments can eliminate the middle management level that essentially suppresses information transfer functions [57]. This loose, task-oriented organizational structure supports knowledge sharing within and between organizations, which also improves the competitiveness of companies. In the majority of responding companies (72.5%), data from enterprise information systems are only available to a limited number of people. And 80% of the companies in the sample believe that a management decision support system with sufficient quantity and quality of data is not necessary to involve lower management in decision making. Centralization-centralization has divided companies: 23.6% of respondents clearly reject the idea of delegating decisions to lower management levels. Only 9.2% of companies have concrete plans to decentralize in the next 1-3 years.

Schemas related to managers have also been highlighted in the mapping and evaluation of decision-making processes. In the majority Hungarian-owned companies, respondents said that managers tend to focus on failures, irregularities and exceptional deviations, and account for all mistakes. In their case, managers delay decisions, wait until things go wrong and problems have to become serious

before they can make a decision and take active action, or even delay a response to urgent issues. This pattern of behavior was negligible in half-owned or majority foreign-owned companies. A similar divergence was also found on the basis of turnover: the second turnover group mentioned above is less focused on failures, in their case managers are not looking for failures and outliers. Interestingly, the larger the company and the more sophisticated the data, the more likely managers are to avoid decisions. Data abundance can lead to both decentralization and higher managerial control. Some companies use data abundance to control, identify and hold to account for errors. For other companies, access to data at lower organizational levels means that decisions are made at the lowest level where the right information is already available. Digitalization can therefore play a very different role in the way companies operate, depending on the management style and organizational culture. Some firms exploit its potential to increase competitiveness, while others use it as a control function, which does not necessarily lead to more efficient operations. The variables of the study (turnover, internationalization, decision making) therefore clearly show management differences, which should be examined in more detail in the future.

Conclusions

The aim of this research for the companies of the future, is to understand what Hungarian business leaders think the company of the future will look like, in order to better understand what factors they consider important for future competitiveness. We were interested in what information they would associate with any future organization, what expectations they would have and by what factors they would describe it. In analyzing the responses, human, cultural and managerial aspects were emphasized over technological or strategic issues. We can therefore assume, that a people-centered approach is associated with the organization of the future.

Analyzing the most frequently mentioned keywords, we can see that the organization of the future will be at the forefront of technology, modern, constantly evolving, digitally mature, innovative and adaptable to the environment. In a rapidly evolving world, there is a growing emphasis on technological development, digitalization, which can be a key pillar of competitiveness, agility and business success.

The turnover categories also showed clear differences and co-movements, which warrant further and more detailed analysis. The pattern anticipated by the models has materialized for some factors: increasingly large companies have more sophisticated systems and routines. However, an unexpected feature was the different behavior of the group of firms in the second revenue category. We hypothesize that this is due to the fact that after the first growth phase, companies move into a "rest" phase and operations, which goes hand in hand with maintaining the status quo and stability, avoiding risks and thus pushing development, digitalization and process innovation to the background.

Internationalization is a key pillar of competitiveness [20], and 51% of the companies in the sample, are currently active in international markets, so identifying the pattern of internationalizing companies and those operating only in the domestic market was an important decomposition factor in the evaluation. The different characteristics of the companies, present on the international market, were clearly visible in the analyses. Companies present on the international market are more open to improvements, support innovation, have a more structured corporate strategy and thus, tend to be more competitive than companies only operating within the Hungarian market.

The perception and experience of the company's manager influences the decisions that determine the company's openness to innovation, development and strategy. The analysis of managerial perceptions will therefore be a priority in the next research phases. This is in line with recent theories [58] that consider the internal set of organizational characters to be more relevant for the future of the firm than external, environmental influences. It seems to be proven [30], that managerial skills and managerial attitudes have an impact on the company's ability to innovate, and thus on the company's strategy and competitiveness. In our sample, managerial attitudes and organizational characteristics and routines were found to be pronounced biasing factors in the information flow, data use and decision-making process. We could detect characteristics that had an impact on corporate digitalization and innovation. These included managerial control, corporate culture associated with error, attitude to risk and characteristics related to the decision process.

The authors aimed to provide a deeper understanding of the competitiveness of Hungarian companies, from the perspective of digitalization and internationalization. The authors' recommendation based on the results are the following:

1. The results show that technological changes, including the challenges of digitalization, are more challenging for larger companies and have the least impact on companies in the second revenue category. Their differentiation can be explained by the fact that they are at different developmental stages, which presents them with very different types of challenges, but they do not yet have the organizational characteristics of classic larger companies. The digitalization performance of companies operating in the domestic market, which are mostly Hungarian-owned, lags behind companies operating in the international market, which are mostly foreign-owned. To make companies more competitive in the future, we suggest to businesses to be more open to innovation and digitalization, to develop employees' skills and suggest management support and risk taking.
2. Companies that are also present in the international market perceive the impact of technology, including digitalization, as stronger and pay more attention to the use of digital technologies, compared to domestic companies that perceive the impact of technological change as less of a challenge. Companies that are

also present on the international market are likely to have the necessary specialized expertise, such as sales knowledge, and thus to perceive expertise as less relevant as a barrier. These companies perform better in improving work process efficiency, identifying new practices and allocating resources accordingly. It is worth exploring networking opportunities to increase the connectivity of companies that think only in terms of domestic markets. This would presumably increase the number and quality of routines learned, which, if incorporated into their own operations, could make domestic firms, which currently have a low level of innovation, more technologically advanced. We advise companies to look for international opportunities in their corporate strategy because internationalization plays a clear role in the development and digitalization of companies through information flows, organizational learning, decision-making processes and corporate culture.

3. The role of managers in the digitalization process was clear from the analysis of the research data. For the managers of companies lagging behind in innovation and digitalization, the role of environmental influences was stronger, their strategy was less formalized and the corporate strategy was less transparent to the members of the organization. Likewise, ownership structure was an important control variable: where there is a single owner, the company is more prone to information flow lock-in, as information is only available to a narrower layer of management, even with higher levels of digitization. In the case of multiple owners and common operational management, digitalization does not only perform a control function, but also plays a role in knowledge sharing and data-driven decision making. Achieving and maintaining a sustainable competitive advantage requires transparent information flows, proactive decision-making and delegation. Digitalization can become a key to the future success of a company if managers change their management style and decision-making mechanisms. To this end, it is necessary to explore in more detail and distinguish the relationship between innovation and firm performance, and to explore whether digitalization has a direct, moderating or mediating effect on competitiveness. This will give us further insights into how digitalization is affecting the professionalization of companies, through their management functions.

Acknowledgement

This survey was organized and supported by Corvinus University of Budapest, Institute of Strategy and Management. Thanks are due to all participants of Future of Organisation research group. Thanks to Prof. László Szerb for making the Small Business Competitiveness Index research database available for a complex analysis.

References

[1] Marciniak, R. (ed.) (2022): The Organisation and Management of the Future. Research Report. Corvinus University of Budapest. DOI: 10.14267/978-963-503-929-6

[2] Huggins, R. – Izushi, H. – Thompson, P. (2013): Regional Competitiveness: Theories and Methodologies for Empirical Analysis. *JCC: The Business and Economics Research Journal,* 6(2), pp. 155-172

[3] Czakó, Zs. – Chikán, A. (2007): Our economic competitiveness from a business perspective – 2004-2006. Budapest Management Review. 38(5), 2-8

[4] Lazíková, J. – Bandlerová, A. – Rohačiková, O. – Schwarz, P. – Rumanovská, L. (2018): Regional Disparities of Small and Medium Enerprises in Slovakia. Acta Polytechnica Hungarica. Vol. 15. No. 8

[5] March, J. G. (2000): Introduction to decision-making. Panem Publishing. Budapest

[6] Pfeffer, J. (2005): Producing Sustainable Competitive advantage through the Effective Management of People. Academy of Management Executive. Vol. 19, No. 4

[7] Hassan M. U. - Shaukat, S. - Nawaz, M. S. - Naz, S. (2013): Effects on innovation types of firm performance: An empirical study on Pakistan's manufacturing sector. Pakistan Journal of Commerce and Social Sciences. Vol. 7, No. 2. pp. 243-262

[8] Szalavetz, A. (2011) Innovation-led growth. *Economic Review*, 58(5), 460-476

[9] Agárdi, I. - Berezvai, Z. - Alt, M.A. (2017): The link between international diversification, innovation and performance in European food retailing. Economic Review. LXIV. 07-08 pp. 805-822

[10] Boikova T. - Zeverte-Rizva, S. - Rivza, P. - Rivza, B. (2021): The Determinants and Effects of Competitiveness: The Role of Digitalization in the European Economies. *Sustainability.* Vol. 2013

[11] Galindo-Martín, M. - Castano-Martínez, M. - Méndez-Picazo, M. T. (2023): Digitalization, entrepreneurship ans competitiveness: an analysis from 19 European countries. *Review of Managerial Science.* Vol. 17. pp. 1809-1826

[12] IMD (2022): https://www.imd.org/centers/wcc/world-competitiveness-center/rankings/world-competitiveness-ranking/

[13] Boda, Gy. (2012): Are we staying on the periphery or catching up? On bottlenecks to Hungarian growth. Seminar paper. Budapest, BCE Institute of Business Economics Competitiveness Research Centre

[14] Csath, M. (2006): Small and medium-sized enterprises as key players in increasing competitiveness. *Reality,* 2006/5

[15] Székely, Cs. (2013): Innovation and creativity. *Economy and Society*. Vol. 5, No. 4. pp. 3-18

[16] Bőgel, Gy. – Krauth, P. (2012): Parallel technological and organisational innovation. *Competitio*. 11(1), pp. 5-19

[17] Hágen, I. Zs. – Holló, E. (2017): The situation of domestic SMEs in terms of competitiveness, innovation and controlling. *Controller Info*. 5 (1)

[18] Gáti, M. – Bauer, A. (2017): A qualitative approach to the organisational interpretation of marketing decisions in SMEs, with a focus on the role of the manager. *Budapest Management Review*. XLVIII. Vol. 2017 12

[19] Rideg, A. (2017): Analysis of the relationship between competitiveness, firm competences and financial performance in the Hungarian SME sector. PhD thesis. PTE GTK Pécs

[20] Szerb L. (2010): Enterprises, entrepreneurship theories, enterprise measurement and the Global Entrepreneurship and Development Index. Academic Doctoral Thesis, Pécs, 2010

[21] Szerb L. – Hornyák M.: A regional comparison of the competitiveness of Hungarian small enterprises. Lengyel I. - Nagy B. (ed.) 2016: Regional competitiveness, smart specialisation and re-industrialisation, JATEPress, Szeged, pp. 307-325

[22] Hámori B. – Szabó K. (2010): An institutional explanation for poor domestic innovation performance. *Economic Review*, LVII Vol. 876-897

[23] Hámori B. – Szabó K. (2012): Innovation and network. *Educatio*. Vol. 27 (2)

[24] Molnár, L. – Udvardi, A. (2016): Competitiveness Yearbook 2016. GKI Economic Research. Budapest

[25] Szalavetz, A. (2004) The information technology revolution and catching-up economies. Budapest, Kossuth

[26] Szerb L. – Komlósi É. – Varga A. (2017): High growth companies Companies in Hungary. *Economic Review*, Vol. LXIV, pp. 476-506

[27] Carmona-Marquez, F. J. – Leal-Millán, A. G. – Vázquez-Sánchez, A.E. (2014): TQM and Business Success Do all the TQM drivers have the same relevance? An empirical study in Spanish firms. *International Journal of Quality & Reliability Management*. Vol. 33, N.3. pp. 361-379

[28] Schlichter, J. – Klyver, K. – Haug, A. (2018): The Moderating Effect of ERP System Complexity on the Growth-Profitability Relationship in Young SMEs. *Journal of Small Business Management.* 59(4) pp. 1-19

[29] Bi, R. - Davison, R. – Smyrnios, K (2019): The Role of Top Management Participation and IT Capability in Developing SMEs Competitive Process Capabilities. *Journal of Small Business Management*. 57(3), pp. 1008-1026

[30] Ónodi, A. – Répáczky, A. (2022): The role of Management in the Success of Innovative Companies. *Budapest Management Review*. LIII (10). pp. 2-14

[31] Incze, E. (2010): The paths of multinationalisation in Hungary - the internationalisation of companies over time. PhD dissertation

[32] Incze, E. (2012): The dynamics of corporate internationalisation - a research model. *Budapest Management Review*, Vol. XLIII, No. 7-8. pp. 32-40

[33] Czakó, E. (2000): Competitiveness at Industry Level - In the Light of Globalization. PhD dissertation, Corvinus University of Budapest

[34] Antalóczy, K. - Sass, M. (2011): Internationalisation of small and medium-sized enterprises - theory and empirics. *Foreign Trade*, Vol. LV. pp. 22-33

[35] Békés, G. - Halpern, L. - Muraközy, B. (2013): Foreign trade and differences between companies. *Foreign Trade.* 60(1), pp. 1-24

[36] Czakó, E. - Juhász, P. - Reszegi, L. (2016): Competitiveness and exports - a comparison of qualitative and quantitative research results at company level. *Budapest Management Review*, 47 (8). pp. 3-14, DOI 10.14267/VEZTUD.2016.08.01

[37] Hurta, H. (2013): Competitive and cooperative management attitudes of small and medium-sized enterprises in Hungary. PhD dissertatioin. SZIE Doctoral School of Business and Management. Gödöllő

[38] Csillag, S. - Csizmadia P. - Hidegh, A. L. - Szászvári, K. (2020): Is the little one nice? Learning and development in small businesses. *Budapest Management Review*, Vol. LI. No. 01

[39] Csákné Filep, J. - Radácsi L. - Timár G. (2020): Factors affecting the survival and growth of Hungarian start-ups. Experiences from expert interviews. *Budapest Management Review*, Vol. LI. No. 01

[40] Chikán, A. – Czakó, E. (2009): In competition with the world. The competitiveness of our businesses on the threshold of the new millennium. Academic Publishing, Budapest

[41] Dunning, J. H. (1993): Multinational Enterprises and the Global Economy. Addison-Wesley, London

[42] Johnson, J. - Vahle, J. E. (1977): The Internationalization Process of the Firm: a model of knowledge development and increasing foreign market commitments. *Journal of International Business Studies*, Vol. 8, No. 1

[43] J. Johanson – Vahle, J. E, (1990): Mechanism of internationalization. *International Marketing Review*, 7, p. 11

[44] Huszák, L. (2020): Growth of SMEs through international market access. Strategic planning for international market entry. Budapest Corvinus University, Textbook chapter

[45] McDougall, P. P. – Oviatt, B. M. – Shrader, R. C. (2003): A Comparasion of International and Domestic New Ventures. *Journal of International Entrepreneurship*, 1(1), pp. 59-82

[46] Oviatt, B. - McDougall, P. T. (1994): A Theory of International New ventures. *Journal of International Business Studies.* 25, 45-64 (1994) https://doi.org/10.1057/palgrave.jibs.8490193

[47] Knight, G. A., Cavusgil, S. T. (2004): „Innovation, organizational capabilities, and the born-global firm", *Journal of International Business Studies* 35(2): 124-141

[48] Stocker, M. - Várkonyi, L. (2022): Impact of market orientation on competitiveness: Analysis of internation- alized medium-sized and large enterprises. *Entrepreneurial Business and Economics Review*, 10(1), 81-95. https://doi.org/10.15678/EBER.2022.100106

[49] Varga, J. (2019): The foundations of organisational competitiveness: How to strengthen corporate competitiveness. Enterprise development in the 21st century, Budapest, http://kgk.uni-obuda.hu/sites/default/files/48_ VargaJanos.pdf

[50] Yang, Y., Deng, L. and Chen, S. (2014) Enterprise Life Cycle, Political Relevance and M & A Strategies. *Management Review*, 26, 152-159

[51] Stocker, M. (2019): Survival, growth, and performance of Hungarian international new ventures. Society and Economy, 41(1), 47-64, https://doi.org/10.1556/204.2019.41.1.4

[52] Dobák, M. - Hortoványi, L. - Szabó, Zs. R. (2012): Conditions for successful growth & innovation. *Budapest Management Review*. Vol. XLIII. No. 12

[53] Karácsony, P. (2020): Analyzing the Relationship between Leadership Style and Corporate Social Responsibility in Hungarian Small and Medium-sized enterprises. *Acta Polytechnica Hungarica*, Vol. 17, No. 7, pp. 183-198

[54] Maxwell, J. A. (2012): Qualitative Research Design: An Interactive Approach, Applied Social Research Method Series. Sage Publications

[55] Almási, A. (2019): The role of organisational factors in the success of the Hungarian SME sector. *Competitio*. Vol. 18, No. 1-2, pp. 108-125

[56] Szanyi-Gyenes, X. – Almási, A. (2021): The role of leadership in the success of small and medium-sized enterprises. *Budapest Management Review*. Vol. 52, No. 3, pp. 40-52

[57] Drucker, P. F. (1988): The Coming of the New Organization. *Harward Business Review*, Vol. 66, No. 1, pp. 45-53

[58] Dörfler, V. (2022): What Every CEO Should Know about AI. Elements in Business Strategy. Cambridge University Press

The Chemical Composition of Post-Consumer Aluminium Scrap – A Challenge in Aluminium Recycling

Ciprian Bulei[1], Imre Kiss[2], Mihai–Paul Todor[1]

[1] Doctoral School
[2.] Department of Engineering & Management
University Politehnica Timișoara, Faculty of Engineering Hunedoara,
5, Revolutiei, 331128 Hunedoara, Romania
ciprian.bulei@student.upt.ro, imre.kiss@fih.upt.ro, mihai.todor@student.upt.ro

Abstract: Aluminium is one of the most recyclable materials, as it can be recycled over and over again, and is one of few materials that keeps its properties after recycling. It can be re-melted and used again and again in new products, making it an environmentally friendly metal and a sustainable material. This makes aluminium an excellent material to meet the needs and challenges of different products. Also, aluminium recycling offers advantages in terms of environmental and economic benefits. Therefore, more aluminium must be collected, sorted, and returned into the economy as new products. Aluminium recycling is the process by which various scrap aluminium can reuse in products after its initial production and involves simply re-melting these scraps. This work provides an overview of the basic aluminium recycling process, using postconsumer scrap in the melting process in few laboratory experiments. Typically, postconsumer aluminium scrap is a mixture of alloys and sometimes even a mixture of metals, the main sources for aluminium scrap being the packaging, technology, construction, and the transport industry. In our experiments, different aluminium scrap sources were considered: mixed packaging aluminium scrap and used beverage can scrap, aluminium from electric cables and aluminium from collected castings. Having in view that the chemical composition is the main challenge in aluminium recycling, mass balance of main aluminium alloying elements is performed. This research provides an overview of the aluminium recycling process, from the scrap upgrading to the melting process.

Keywords: aluminium recycling; post-consumer aluminium scrap; re-melting; chemical composition

1 Introduction

Aluminium became the most important structural material of the 20th and 21st Centuries, currently, aluminium being the most widely used non-ferrous metal in

the world. The use of aluminium and its alloys has begun to experience a continuous development in the various existing industries and represents one of the most important categories of materials used in modern technology [1-4]. Aluminium is used in many industries for the manufacture of millions of different products and is very important for the world economy. Aluminium structural components are vital for the aerospace industry and very important in other areas of transport and construction where ease, durability and strength are required. The largest amount of aluminium is used to obtain its alloys, which is widely used in various techniques. Without going into technical details now, very many modern industries depend on this metal [1-4]. There is no branch of industry that does not use aluminium and its alloys in the form of castings, forgings, moulds, sheet metal, strips, foils, wire, profiles, engine blocks, etc. Basically, aluminium offers smart and practical solutions to modern life. Without this metal, today's society is unimaginable. Thus, it has practically become a symbol of progress.

The considerable expansion, in recent years, of the widespread use of aluminium and its alloys, processed in different forms, in areas such as automobiles, construction, packaging, electricity, household appliances, has led to a significant increase and accumulation of the significant amount of waste [5-13]. As a result, a special branch of metallurgy and industry was created, called the industry of secondary aluminium or recycling of aluminium from waste. The importance of recycling aluminium results from the analysis of the effects that secondary aluminium production has on the economy and the environment. Aluminium offers a lower-weight alternative to steel and it fits greatly into a circular economy since it is highly recovered and reused in new products (Fig. 1) [1, 6-11].

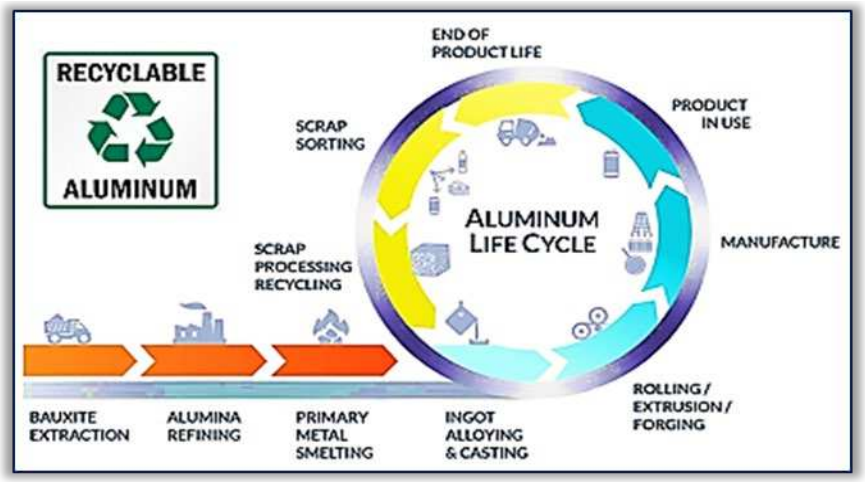


Figure 1
The aluminium life cycle

Among the general objectives of waste management are [1-4]:

— intensifying concerns about reducing the amount of waste generated;

— exploiting all the technical and economic possibilities for the recovery and recycling of waste in order to reduce the amount of waste disposed of.

— awareness of the factors involved regarding the need for separate collection;

When setting the objectives of waste management, the following aspects must be taken into account [1, 4-9]:

— not all used goods are completely reintroduced into the economic circuit;

— the reuse of some waste involves a high energy consumption;

— waste recycling is justified only when profitability and ecological balance are favourable;

— the recycling of some waste is limited by technological barriers;

— the existence of a functional market for the products obtained from waste recycling.

And because the aluminium:

— is light, durable, versatile and, above all, it is very easy to recycle – it can melt easily and quickly turn into a new product;

— is one of the few materials for which recycling costs are exceeded by the selling price of the recycled product – as are steel and copper;

— provides a rare combination of valuable properties: corrosion resistance, high strength and low-density;

— has unique recycling qualities: can be recycled repeatedly without losing quality, is 100% recyclable and can be re-valued indefinitely (in fact, aluminium can be recycled endlessly without loss of material properties);

— recycled, saves energy: aluminium production is a complex process that involves a large amount of energy, recycling aluminium, in turn, needs only 5% of this energy – the re-melting of the used aluminium saving up to 95% of the initial energy needed to produce raw material;

— uses less energy for recycling and recycling is self-sustaining due to the high value of the aluminium used, proving to be economical;

— is still in use, in proportion of 75% of the amount of aluminium produced in time, 60% being still in the first use;

— is extremely versatile and is a material with many uses in a multitude of different products, from cars to window frames and from airplanes to packaging,

aluminium waste is an important resource that is extremely valuable. For this reason, the aluminium used is rarely lost. The aluminium industry has every interest in promoting aluminium recycling as part of its industrial strategy [1-13].

Due to its unique properties (low weight, malleability, conductivity, corrosion resistance and impermeability) and its ability to be recycled infinitely, aluminium is a raw material with multiple options for use [1-13]. Recycled aluminium is part of the solution. Given the current energy problems, the increasing price of oil, the problem of water resources but also the major risks on the environment, it is easy to see how useful the non-ferrous recycling activity is. Compared to other high-volume materials, such as copper, zinc, magnesium, and steel, aluminium production has one of the widest energy differences between the primary and secondary routes.

2 Materials

The material valorisation of aluminium waste includes any operation or succession of dismantling operations, sorting, cutting, shredding, pressing, baling, melting – casting performed on a waste by industrial processes, in order to transform it into secondary raw material. Material valorisation involves the substitution of some raw materials, the waste being used again due to its material characteristics. From the point of view of the recovery possibilities, the waste is classified into [1-4]:

— recirculation waste, if it is recovered on the spot in the foundry of the enterprise that produced it;

— tradable waste, when collected for proper processing and recovery;

Depending on the time after which it returns to the circuit as waste and the area from which it originates, the waste can be classified into [1, 2, 12]:

— pre-consumer waste (manufacturing waste or "new waste"), is typically scrap that comes from the production process and has never been used before it is re–melted again (Fig. 2, left);

— post-consumption waste ("old waste"), is aluminium that has already lived a "life" and has now been collected, sorted and recycled to be turned into something new product (Fig. 2, right).

Thus, the great potential for preventing the generation of aluminium waste is effective in two stages [1-3]:

— in the production phase (Fig. 2, left);

— at the end of a product's life (after consumption – Fig. 2, right).

Figure 2
Aluminium waste in the production phase respectively at the end of a product's life (after consumption) – example for the beverage cans

Aluminium scrap is often categorized as "new scrap" and "old scrap". "New scrap" originates during the manufacturing of semi-fabricated and final products (shavings, off-cuts, casting parts, etc.) where the quality and chemical composition are usually known. Usually, the "new scrap" category is molten without any preliminary treatment. "Old scrap" refers to those products collected after disposal from consumers, thus at the end of their life (cables and electric wires, castings, auto parts, window frame, beverage cans, etc.). This raw material is more contaminated than "new scrap" and preliminary treatments of the scrap are generally necessary [1-3,14-18]. Therefore, the aluminium destined to material valorisation may be separated into two categories, namely:

— new waste: represents the technological waste generated by the processes of casting, thermal or mechanical processing (casting networks, scrap, casting burrs, ends resulting from the cutting of extruded products, edges from cutting sheets, sheet ends, scraps from stamping, spirals, chippings, etc.) and materials with residual aluminium content (such as slags discharged during elaboration operations).

— old waste: are those that come from products or articles put out of service at the end of their life cycle. Such wastes are packaging, those that come from transports, components of vehicles, construction carpentry, old electrical conductors, household appliances, etc. The old waste comes from the collection units, equipped with shredders, magnetic separators, and separation installations in dense environments.

Used aluminium mainly comes from two channels, namely, waste from domestic and industrial consumption. The way in which recycling is carried out is fundamental, its classification plays an important role in obtaining a material that has the same use as before its recovery or is intended for another type of products.

3 Methods

The aluminium recycling circuit can be achieved through two streams, one closed and the other open. In the closed loop, easily recoverable aluminium waste, which has the same provenance and known chemical compositions, can be recycled. They are recycled through smelting operations, being usually used to manufacture the same products (an example being that of car wheels). After the open loop, waste mixtures are most often processed [1-12]. They have variable chemical compositions, require additional refinements and alloys, so that the alloys obtained from their re-melting can be used to make new products, most often by casting.

The streams that process aluminium waste to obtain secondary aluminium or aluminium alloys are composed of two main steps:

— preparation (or pre-treatment) of waste: sorting, dimensional preparation, removal of impurities (purification);

— elaboration with refining of aluminium or aluminium alloys: melting, refining, alloying, casting.

In addition to these stages (Fig. 3), the flow may include other additional stages / operations, depending on the quality of the waste, the sources of waste, the endowment with auxiliary equipment, constructively modified furnaces, the specifications of the products obtained [1-12].

The heart of aluminium recycling is the melting aggregate, in which the waste melting operations are performed in order to obtain new raw materials. Melting is the process by which aluminium scraps can be reused in products after their initial production.

Figure 3
The stream of aluminium waste to recycling into new products

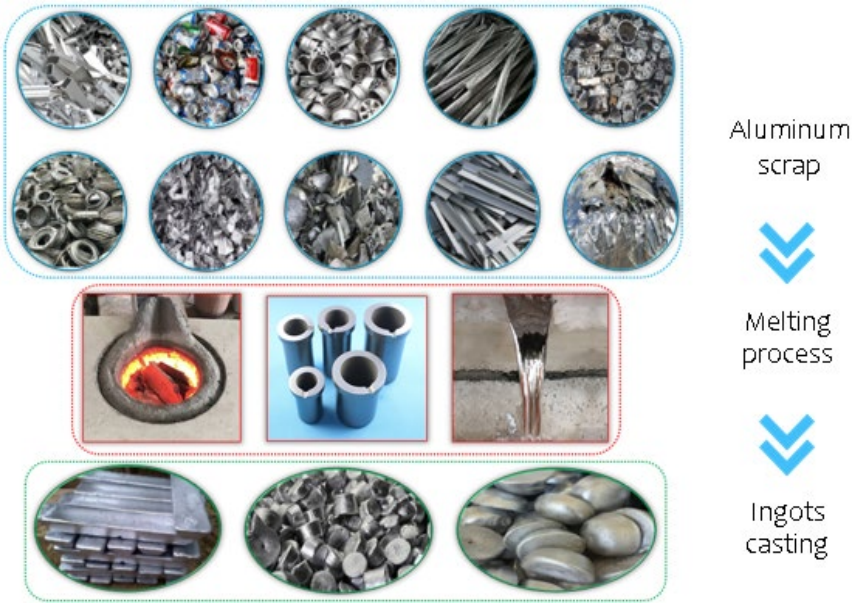


Figure 4
Re-melting the aluminium scrap

The process (Fig. 4) simply involves re-melting metals, which is much less expensive and more energy-intensive than creating a new aluminium by electrolysis of aluminium oxide (Al2O3), which must first be extracted from bauxite ore and then refined using the Bayer process. The selection of the melting furnace is a critical aspect and it depends on the quality and quantity of the scrap.

4 Results & Discussions

The aluminium recycling circuit can be achieved through two streams, one closed and the other open. In the closed loop can be recycled easily recoverable aluminium waste, which has the same provenance and known chemical compositions. They are recycled through melting process (Fig. 5) and are usually used to manufacture the same products (an example being that of car rims or beverage cans). After the open loop, waste mixtures are most often processed. They have variable chemical compositions, require additional refinements and alloying, so that the alloys obtained from their re-melting can be used to make new products, most often by casting.

Figure 5
Aluminium scraps melting process and cast ingots

At the moment the aluminium can (Fig. 6a) is one of the most recycled packaging in the world. Many of the food and beverages that we buy every day are packed in aluminium or steel cans, both materials can be recycled in order to manufacture new containers or other metal products. Sooner or later, each can ends up being thrown in the trash, where it begins its journey to recycling. However, many cans end up in landfills. The melted aluminium cans chemical composition in Table 1 is presented.

Figure 6
Aluminium scraps: beverage cans, castings and electric wires

Aluminium from the car rims (Fig. 6b), engine blocks or membranes, castings, etc. are intended for the manufacture of ingots that present the highest percentage of secondary aluminium production, about 70% of the total production. Therefore, the environmental and economic impact on the use of primary aluminium can be significantly reduced by recycling it. The melted aluminium castings chemical composition in Table 2 is presented.

Table 1
Aluminium cans – Chemical composition (%)

Al	Mg	Fe	Si	Mn	Sn	Pb
97.83	0.402	0.567	0.281	0.284	0.0077	0.337
Ga	Na	Cr	Ni	Cu	Sb	K
0.0112	0.0265	0.0091	0.0053	0.130	0.0107	0.0012
Ca	Zn	La	Zr	Mo	Ti	Ar
0.0391	0.0349	0.0038	0.0021	0.0012	0.0139	0.002

Table 2
Aluminium castings – Chemical composition (%)

Al	Mg	Fe	Si	Mn	Sn	Pb	Zr
95.18	0.739	1.65	0.280	0.725	0.0374	0.980	0.0036
Ga	Na	Cr	Ni	Cu	Sb	K	Ti
0.0119	0.0216	0.0185	0.0107	0.160	0.0575	0.0014	0.0243
Ca	Zn	La	P	Mo	V	As	
0.0425	0.0379	0.0030	0.0031	0.0027	0.0077	0.0022	

Aluminium from electric cables (Fig. 6c) is very pure, and for this fact its applications after recycling are extensive, being used in the manufacture of high purity alloys or as mixtures to reduce the percentage of alloying present in another recovered aluminium alloy. The melted aluminium wires chemical composition in Table 3 is presented.

Table 3
Aluminium wires – Chemical composition (%)

Al	Mg	Fe	Si	Mn
99.55	0.135	0.125	0.0881	0.0291
Ga	Na	Ce	Ni	Cu
0.0175	0.0127	0.0086	0.0071	0.0059
Ca	Zn	La	Hf	Mo
0.0051	0.0036	0.0022	0.0019	0.0014

Different aluminium scrap sources were considered: mixed packaging aluminium scrap and used beverage can scrap, aluminium from electric cables and aluminium from collected castings (Fig. 7). Copper (Cu), silicon (Si), magnesium (Mg), zinc (Zn), manganese (Mn), nickel (Ni) are used for aluminum to set up the main alloys. Along with them are used elements that improve some aspects of the performance of the alloys, known as corrective. Small percentages of titanium (Ti), zirconium (Zr), chromium (Cr), bismuth (Bi), lead (Pb), and also tin (Sn) and iron (Fe) are added for special purposes, the latter of which is always present as an impurity.

Figure 7
Aluminium scraps: melted aluminium and cast aluminium ingots

The quality of the metal is defined by three characteristics, namely:

- control of trace elements;
- eliminating non-metallic inclusions, and
- reduction of dissolved gas emissions.

If these inclusions are not eliminated from the melt before the material solidification, they lead to the worsening of the mechanical properties, such as: tensile strength, fragility, corrosion resistance, thermal and elective conductivity, etc. With all the measures that are taken in the practice of elaboration of aluminium alloys in liquid state, they still contain a certain amount of metallic impurities, non–metallic and gaseous inclusions.

The effects of metallic impurities present in aluminium alloys can be negative or positive, these effects depend on their quantity and combination in the melt. Usually the impurities existing in aluminium alloys after, are: Cu, Fe, Si, Zn, Na, Mg, Ti, V, Cr, Mn, Zr, etc. Impurities in technical aluminium may be classified in the following categories:

- impurities that react chemically, forming various compounds, easily fusible (Fe, Si);
- impurities that form easily fusible eutectics, but which are practically insoluble in the solid state (Sn, Pb, Bi);
- impurities that form chemical compounds that are difficult to fuse, but insoluble in the solid aluminium (As, Sb, Se, Te);
- partially soluble impurities in aluminium in the solid state (Si, Cu, Mg, Zn, etc.);
- impurities from the gas phase (hydrogen, oxygen, nitrogen).

The mechanisms by which these metals act negatively or positively in aluminium and its alloys are those that lead to the formation of chemical and eutectic compounds to the solidification of the material. The most common inclusions in aluminium alloys are the oxide particles that are found on the surface of the load,

and then end up in the metal bath. In some cases non-metallic inclusions are formed in the molten alloy, and in others they already pre-exist in it and participate in subsequent processes. As a result of this, we can classify these inclusions into two categories, namely:

— non-metallic inclusions due to metal, which are also called endogenous inclusions, and
— non-metallic solid inclusions coming from the outside (from the shape material, the oven lining, the fondants used, the impurities in the load, etc., which are also called exogenous inclusions.

The need for improved mechanical properties and new applications has led to the continuous development of new kinds of aluminium alloys with specific chemical composition. To remove impure elements from a molten bath is impractical or inconvenient. As a result, the scrap is usually recycled, which avoids the refinement stage. A sustainable solution is the improvement of the efficiency of aluminium recycling in the production chain, which includes the scraps re–melting process, after a correct sorting process.

The chemical composition of the alloys is strictly related to the scrap quality. Therefore, recycled aluminium presents a certain amount of impurities, generally not present in primary alloys, and the alloying elements are more difficult to manage. The melting phase is critically analysed in terms of technological evolution and furnace selection, as it is the most important choice to optimize the melting rate. Fluxing and slag treatments have also been considered to complete the production chain.

During the multiple recycling, more and more alloying elements are introduced into the metal cycle. This effect is put to good use in the production of casting aluminium alloys, which generally need these elements to attain the desired alloy properties.

Conclusions

All aluminium products can be recycled after use. When decommissioned, practically all aluminium products have some value that guarantees that it is possible to create new values by recycling them into products. Since aluminium can be recycled almost without loss of quality, and because it stores a high intrinsic value, there is a strong natural incentive to recover and recycle aluminium products after they are decommissioned. The recycling yield is dependent on the lifetime of products made from aluminium. Average service life-spans range from a few months for the packaging segments, 10-12 years for car castings, to more than 30-40 years for construction and electricity.

Advantages of recycling aluminium waste:

— conservation of natural resources;
— reduction of storage space;
— protecting the environment;

Whether measured in terms of quantity or value, the use of aluminium surpasses that of all other metals except iron, and is important in virtually all segments of the world economy. Pure aluminium is soft and weak, but it can form alloys with small amounts of copper, magnesium, manganese, silicon and other elements that have a wide range of useful properties.

Recovering this metal from waste (through recycling) has become an important part of the aluminium industry. Sources of recycling of aluminium include cars and window frames, appliances, containers and other pr oducts. Recycling is very convenient. The benefits of recycling are undeniable. Each waste selectively collected and recycled/reused helps to save natural resources, reduce pollution and increase the quality of life.

The market exploitation of aluminium waste is worth considering and is economically viable. The use of secondary aluminium provides not only huge resource savings, but also an environmental benefit, with the reduction of emissions compared to the electrolytic process and the guarantee that the material will re-enter a production cycle, therefore, without the risk of an ecological impact. The advantages of this behaviour of this light metal are clear: secondary aluminium is the equivalent of primary aluminium, even after several life cycles. A secondary route for aluminium production is available using aluminium scrap and recycling. It is claimed that recycling saves resources, decreases the need for landfill space and, in the case of non-renewable resources, such as metals, and prolongs the necessary period to deplete them.

Thanks to the immense availability of recycled material (landfilled and collected) and thanks to its mechanical properties the aluminium will be the material of the future. That's why we must focusing on recycled aluminum from various sources. And the recycled aluminium will be destined mostly for the most important sectors that use this material: automotive and construction (cast and extruded aluminum). But, the excellent recyclability of aluminium, together with its high scrap value and low energy needs during recycling make it highly desirable to all industries, in air, road and sea transport (aircraft, automobiles, bicycles, boats), packaging (beverage cans to food and drinks containers, aerosols, tubes and foil), electronics (phones, computers and its components) and electrical power transmission (wire), kitchen products (aluminium cookware and containers), sports and recreation (skates, instruments, and yacht fittings), and composites industry (as matrix in lightweight composites).

The aluminium industry aims to ensure that all types of waste generated are of high quality and that they are minimised and recycled in the most efficient way. Aluminium is practically a permanent material, a material whose inherent properties do not change during use and following repeated recycling in new products. However, it depends on the recycling method to what extent the re-melted aluminium is reintroduced into the circuit. In fact, recycling is a joint effort and the circular economy is a lifestyle, in which waste is treated as a resource and

turned into valuable raw materials that can be reused for new products. Millions of tonnes of recyclable waste are selectively collected annually around the world, with recycling enjoying strong demand from industry and construction companies, as well as constant demand from households. Sources for aluminium scrap include packaging (beverage cans), construction (window frames), energy (electric wires) and the transport industry (auto parts). The greatest opportunity to increase the availability and quality of scrap is in consumer packaging and the automotive sector.

Acknowledgment

These studies was carried out on the basis of extensive research over the last 5 years, undertaken within the Faculty of Engineering Hunedoara. These researches are part of a complex research program, included in a doctoral study plan of University Politehnica Timisoara, which analyzes different possibilities of recycle various landscaped and collected aluminium scraps.

References

[1] D. Raabe, D. Ponge, P. J. Uggowitzer, M. Roscher, M. Paolantonio, C. Liu, H. Antrekowitsch, E. Kozeschnik, D. Seidmann, B. Gault, F. De Geuser, A. Deschamps, C. Hutchinson, C. Liu, Z. Li, P. Prangnell, J. Robson, P. Shanthraj, S. Vakili, C. Sinclair, L. Bourgeois, S. Pogatscher, Making sustainable aluminum by recycling scrap: The science of "dirty" alloys, *Progress in Materials Science*, 128, (2022) https://doi.org/10.1016/j.pmatsci.2022.100947

[2] S. Capuzzi, G. Timelli, Preparation and melting of scrap in aluminum recycling: A review. *Metals (Basel)* (2018) 8:249, https://doi.org/10.3390/met8040249

[3] A. T. Tabereaux, R. D. Peterson, Chapter 2.5 – Aluminum production, treatise on process metallurgy, *Elsevier*, 2014, 839-917, https://doi.org/10.1016/B978–0–08–096988–6.00023–7

[4] P.de Schrynmakers, Life cycle thinking in the aluminium industry. *The International Journal of Life Cycle Assessment*, 14, 2-5 (2009) https://doi.org/10.1007/s11367–009–0072–x

[5] G. Olivieri, A. Romani, P. Neri, Environmental and economic analysis of aluminium recycling through life cycle assessment, *International Journal of Sustainable Development & World Ecology*, 13:4, (2006) 269-276, https://doi:10.1080/13504500609469678

[6] M. Niero, S. I. Olsen, Circular economy: To be or not to be in a closed product loop? A Life Cycle Assessment of aluminium cans with inclusion of alloying elements, *Resources, Conservation and Recycling,* 114, (2016) 18-31, https://doi.org/10.1016/j.resconrec.2016.06.023

[7] G. Gaustad, E. Olivetti, R. Kirchain, Improving aluminum recycling: A survey of sorting and impurity removal technologies. *Resources,*

Conservation and Recycling, 58, (2012) 79-87, https://doi.org/10.1016/j.resconrec.2011.10.010

[8] G. Gaustad, E. Olivetti, R. Kirchain, Improving aluminum recycling: A survey of sorting and impurity removal technologies, *Resources, Conservation and Recycling*, 58, (2012) 79-87, https://doi.org/10.1016/j.resconrec.2011.10.010

[9] S. K. Das, Emerging trends in aluminum recycling: Reasons and responses. *Light Metals*, 4, (2006) 911-916

[10] J. R. Duflou, A. Erman Tekkaya, M. Haase, T. Welo, K. Vanmeensel, K. Kellens, W. Dewulf, D. Paraskevas, Environmental assessment of solid state recycling routes for aluminium alloys: Can solid state processes significantly reduce the environmental impact of aluminium recycling?, *CIRP Annals*, 64(1), (2015) 37-40, https://doi.org/10.1016/j.cirp.2015.04.051

[11] V. K. Soo, J. Peeters, D. Paraskevas, P. Compston, M. Doolan, J. R. Duflou, Sustainable aluminium recycling of end-of-life products: A joining techniques perspective, *Journal of Cleaner Production*, 178 (2018) 119-132, https://doi.org/10.1016/j.jclepro.2017.12.235

[12] S. Eggen, K. Sandaunet, L. Kolbeinsen, A. Kvithyld, Recycling of aluminium from mixed household waste. *Light Metals* (2020) 1091-1100, https://doi.org/10.1007/978–3–030–36408–3_148

[13] S. Shamsudin, M. A. Lajis, Z. W. Zhong, Evolutionary in solid state recycling techniques of aluminium: a review. *Procedia CIRP*, 40, (2016) 256-261, https://doi.org/10.1016/j.procir.2016.01.117

[14] C. Bulei, M. P. Todor, T. Heput, I. Kiss (2018) *Recovering aluminium for recycling in reusable backyard foundry that melts aluminium cans*. In IOP Conference Series: Materials Science and Engineering, 416(1), https://doi:10.1088/1757–899X/416/1/012099

[15] C. Bulei, M. P. Todor, I. Kiss (2018) *Metal matrix composites processing techniques using recycled aluminium alloy*. In IOP Conference Series: Materials Science and Engineering, 393(1), https://doi:10.1088/1757–899X/393/1/012089

[16] C. Bulei, M. P. Todor, I. Kiss, V. Alexa (2019) *Aluminium matrix composites: Processing of matrix from re-melted aluminium wastes in micro-melting station*. In IOP Conference Series: Materials Science and Engineering, 477(1), https://doi:10.1088/1757–899X/477/1/012049

[17] C Bulei, I Kiss, V Alexa, Development of metal matrix composites using recycled secondary raw materials from aluminium wastes, *Materials Today: Proceedings*, 45(5), 2021, pp. 4143-4149, https://doi.org/10.1016/j.matpr.2020.11.926

[18] U. M. J. Boin, M Bertram, Melting standardized aluminum scrap: A mass balance model for Europe. *Jom*, 57(8), (2005) 26-33

Analyzing the Impact of Short-Term Cyclic Thermal Ageing on PVC Insulated Low Voltage Samples with Polarization/Depolarization Current Measurement

Semih Bal, Zoltan Adam Tamus

Department of Electric Power Engineering, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics
Egry József u. 18, H-1111 Budapest, Hungary
E-mail: semihbal@edu.bme.hu; tamus.adam@vik.bme.hu

Abstract: Cables are critical elements of the network. Therefore, their integrity is of utmost importance. Many factors influence the insulation integrity of the cables, such as manufacturing procedures, operating conditions, and various stresses, during their lifetime. Increasing the number of renewables resulted in increased distributed generation, which may cause reverse power flow. The electricity demand is constantly growing with the increased number of novel appliances on the consumer side, electric vehicles, etc. This fact can create short-term overloads, which may cause a temperature elevation beyond the cable's maximum operating limit, compromising insulation degradation. Therefore, this research investigates the impact of short-term cyclic thermal ageing on SZRMtKVM-J 4 x 6 mm PVC insulated low voltage cable using the polarization/depolarization current measurement (PDC). Accelerated thermal ageing tests were performed by placing the samples inside the temperature-controlled oven for 6 hours/round. After each round, the polarization and depolarization currents were measured. DIRANA Dielectric Response Analyzer by Omicron has been employed for the measurements. The results have shown an increase in the polarization and depolarization current.

Keywords: PVC Insulation; Thermal Ageing; PDC Measurement; Polarization/ Depolarization Current

1 Introduction

PVC insulated low voltage (LV) cables are commonly used in several applications, like electric distribution, power plants, communication, security systems, etc. [1] [4]. The 2030 Climate Target Plan encourages countries to minimize greenhouse gas emissions by at least 55% by 2030, aiming at achieving climate neutrality by 2050 [5]. Therefore, countries are abandoning fossil fuels to reduce greenhouse gas

emissions, meaning more share for renewables. The term smart grid is getting more attention as we move towards the decentralized generation with increased renewables. The distributed generation means the current can flow in both directions, from load to the grid and vice versa. The new appliances connected to the network can cause short-term overloads. These overloads may increase the temperature above the operation limit, which leads the insulation degradation [4], [6], [7]. Therefore, the impact of cyclic short-term thermal ageing has been the focus of this research.

Variations of PV power generation can be modelled in several ways [8]. This paper employs a similar method by utilizing a constant ageing temperature on cable samples in order to simplify the laboratory simulation of accelerated ageing [9]. The purpose is to model the peak power generation of photovoltaic (PV) systems that typically occurs around noon when production is at its highest.

It is vital for the system's safety and reliability; the insulation should perform its intended function, preventing current from flowing [10]. The cables are faced with some stresses in normal operation, which causes a mutation in the molecule structure and degrade the insulation. This is known as insulation ageing [11]-[14]. The electricity market is growing in a direction in which several independent companies compete, such as transmission operators, system operators, and distribution operators. These companies must cut costs while keeping the supply up and running. Costs will not be reduced with the "unscheduled maintenance" philosophy. As the name implies, unscheduled maintenance occurs only when necessary, such as when a transformer or cable fails. However, this usually interrupts the energy supply, making breakdowns much more expensive than preventative maintenance [15]. Therefore, the condition monitoring (CM) of the cables in the network is essential.

Several condition-monitoring techniques are employed by researchers. These CM techniques can be destructive and non-destructive [16]. Since most LV distribution cables are buried underground, a non-destructive CM technique has been used in this study.

Polarization-Depolarization (PDC) measurement is a well-known condition monitoring technique based on the dielectric relaxation for the transformers and cables [17]. It is based on the measurement of polarization (charging, absorption) current and the depolarization (discharging, desorption) current. Morsalin et al. stated that the time constant of polarization/depolarization currents correlates well with ageing and insulation deterioration [18].

This work presents a novel approach to studying the behaviour of insulating materials exposed to short-term, cyclic thermal stress. It is known that each material responds in distinct way when subjected to these stresses [19]. To assess the changes in conductivity, a PDC measurement technique is employed on a specific type of cable (SZRMtKVM-J 0.61/1 kV 4x6 mm^2 PVC insulated) after each thermal ageing cycle. DIRANA Dielectric Response Analyzer, commonly used for condition

monitoring of power and instrumental transformers, is utilized as a measuring device for this low-voltage (LV) cable type. This research extends the application of the DIRANA and provides valuable insights into the unique responses and conductivity changes exhibited by insulating materials under short-term cyclic thermal stress conditions.

This paper presents an experimental study of PVC insulated LV cable exposed to cyclic thermal stress. Following the introduction section, a brief explanation of PDC technique is given. The experimental work is given in the third section, where a detailed introduction about the cable sample, ageing temperature and how it was chosen, the measurement setup, and accelerated ageing tests are presented. The fourth chapter is dedicated to the results and discussions. Last, but not least, the fifth chapter is the conclusion. It is worth mentioning that this is an ongoing study, and the preliminary results are presented here. The dielectric relaxation measurement was applied on thermally aged cable samples in the time domain, and polarization/depolarization currents were measured.

2 Methodology

Researchers widely accept Dielectric Response Measurement (DRM) in their study to investigate the insulation's dielectric characteristics like tanδ, capacitance, insulation resistance, polarization, and depolarization current [20]. It is possible to conduct DRM in the time domain or frequency domain. These dielectric measurements are non-destructive, which is quite advantageous because these techniques can make it possible to measure the dielectric parameters without removing the samples from their network.

The PDC measurement is based on two processes. These are charging and discharging the test sample. The circuit diagram of PDC measurement is shown in Figure 1. The sample is charged with a voltage source (V_{ch}) for a definite time through a closed S_1 switch. The flowing current is called the polarization current (i_p) during charging—the polarization current's amplitude changes by orders of magnitude over time. I_p is dependent on the insulation conductivity. Once the charging period is over, S_1 opens and S_2 closes to discharge the sample. The depolarization current (i_d) is measured. It is shown in Figure 1 that i_p and i_d flow in opposite directions.

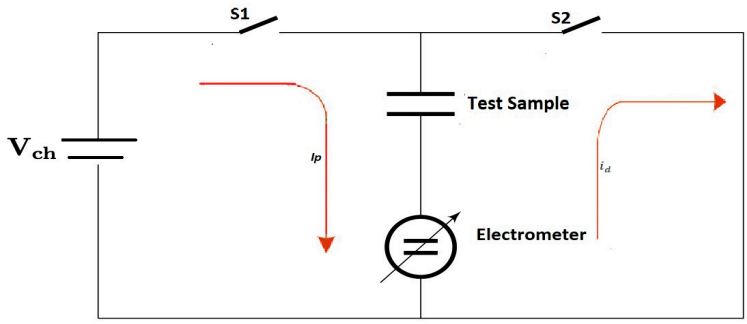


Figure 1
The circuit diagram of PDC

The polarization current is explained in Equation 1:

$$i_p(t) = C_0 U_c \left[\frac{\sigma}{\varepsilon_0} + f(t)\right] \quad (1)$$

Where:

C_0 is the geometric capacitance, σ is conductivity and *f(t)* is the dielectric response function.

The depolarization current is explained in Equation 2;

$$i_d(t) = C_0 U_c [f(t) - f(t + t_c)] \quad (2)$$

If the polarization process is sufficiently long enough, conductivity σ can be expressed as in Equation 3.

$$\sigma = \frac{\varepsilon_0}{C_0 U_c}[i_p(t) - i_d(t)] \quad (3)$$

Figure 2 depicts the timing diagram of PDC measurement. The polarization and depolarization currents are both affected by the insulation geometry and material properties [21] [24].

The polarization and depolarization currents are influenced by the properties of the insulation material and also by the geometric structure of the insulating system [25-26].

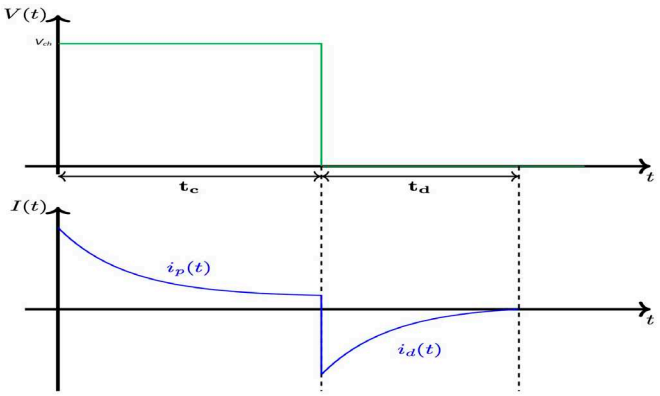


Figure 2
The timing diagram of PDC

3 Experiment

3.1 Cable Sample

In this experiment, 0.6/1 kV 4 x 6 mm2 PVC insulated cable (type SZRMtKVM-J) samples, made by Pyrismian Hungary Kft., were used. Three cable specimens, each of 50 cm, were cut from the bulk of the cable. The cable samples meet the requirements of IEC 60502-1, which specifies the standards for power cables with solid insulation in terms of design, dimension and testing. According to the cable datasheet, the maximum operating temperature is 70°C. The cable consists of five layers, which is illustrated in Figure 3.

Figure 3
The cable structure

The layers from inside to outside as below:

1. 4 x copper conductor.
2. PVC core insulation.
3. PVC tape belt.
4. Steel armour.
5. PVC jacket.

3.2 Thermal Ageing

As mentioned, the cable temperature may rise beyond the maximum limit due to short-term overloads. Therefore, this paper aims to observe the effect of these short-term temperature rises. The accelerated thermal ageing was performed by placing the samples in the temperature-controlled oven for a short time and was repeated over the cycles. The ageing temperature was set to 110°C.

IEC 60502-1 states that thermal ageing can be performed at least 10±2 °C higher than the maximum conductor temperature [27]. Figure 4, taken from IEC 60502-1, shows that the maximum conductor temperature of PVC insulated cable can be 70 °C in normal operation and 160 °C in short-circuit operation.

Table 3 – Maximum conductor temperatures for different types of insulating compound

Insulating compound		Maximum conductor temperature °C	
		Normal operation	Short-circuit (5 s maximum duration)
Polyvinyl chloride	(PVC/A)		
	Conductor cross-section ≤300 mm ^{2}	70	160
	Conductor cross-section >300 mm ^{2}	70	140
Cross-linked polyethylene	(XLPE)	90	250
Ethylene propylene rubber	(EPR and HEPR)	90	250

Figure 4

The maximum conductor temperature [27]

By choosing 110 °C as the ageing temperature, it is desired to apply higher thermal stress to the samples than given in the standard. Also, it is aimed to make the comparison more effortless with the previous studies [4], [6], [7], [28], [29]. Eighteen hours, 6 hours/round, ageing time have been reached. Equivalent ageing time can be calculated using the Arrhenius equation, which is Equation 4.

$$\frac{t_s}{t_a} = e^{\frac{E_a}{k}\left(\frac{1}{T_s}-\frac{1}{T_a}\right)} \tag{4}$$

Where:

- t_s= Operating time.
- t_a=Equivalent ageing time.
- E_a=Activation energy, taken from the literature and set to 80kJ/mol (0.829eV) [30].
- k=Boltzman constant: 8.617333262×10^{-5} eV/K.
- T_s=Absolute temperature (70°C) in Kelvin: 343.15 K.
- T_a=Absolute ageing temperature (110°C) in Kelvin: 383.15 K.

It can be calculated that eighteen hours of ageing at 110°C would equal 330 hours at 70°C. Table 1 shows the equivalent ageing times after each round.

Table 1
Equivalent ageing time

Round	Ageing time (hours) (t_a)	Equivalent time in operation (hours) (t_s)
1	6	110
2	12	220
3	18	330

3.3 PDC Measurement

In the time domain, i_p and i_d currents were measured using DIRANA. The general dielectric configuration was chosen from DIRANA library. This configuration can be seen in Figure 5.

The parameters of the measurement were set as follows:

- Test voltage: 200 V.
- Polarization time: 5000 s.
- Depolarization time: 1000 s.

Zaengl et al. suggested that the polarization time should be at least five times higher than the depolarization time [15]. Therefore, the polarization and depolarization times were chosen accordingly. The reason for choosing 1000 s for discharging time is to investigate the slow polarization process. Faraday cage was used to avoid external disturbances. The measurement was performed inside the laboratory, where the temperature was constant at 24±0.5°C.

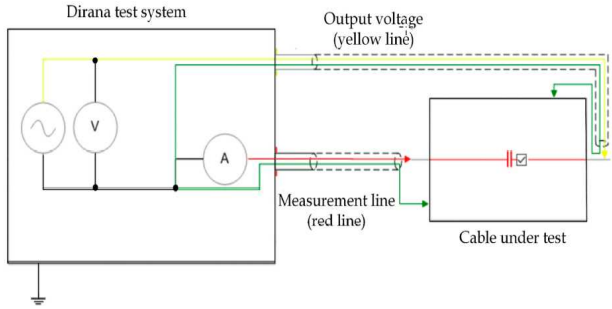


Figure 5
General dielectric measurement configuration from DIRANA library

i_p and i_d current for the core and jacket were measured. 30 cm length of aluminium foil is used to create a conductive surface by wrapping the jacket with it.

For core measurement;

- The measuring probe is connected to the conductor of the cable.

- The other cores and the aluminium belt were short-circuited, and the output probe is connected to them.
- The guard probes are connected to the core insulation, which is being measured.
- The ground probe is connected to the earthing point of the Faraday cage.

For jacket measurement;

- The measuring probe is connected to the jacket and wrapped with aluminium foil.
- All core conductors and aluminium belt were short-circuited, and the output probe is connected to them.
- The guard probes are connected to the jacket.
- The ground probe is connected to the earthing point of the Faraday cage.

The guard probes are used to eliminate the surface leakage current. A single-point grounding was used to prevent any circulating current that could interfere with the results. The sample measurement arrangement is shown in Figure 6.

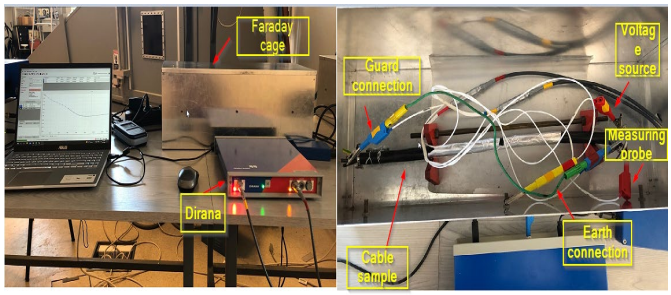


Figure 6
PDC measurement arrangement – Core measurement

The cable samples were short-circuited to the ground between each measurement round to ensure that no remaining charges existed to affect the results. The polarization current curves show a similar trend after each round. This proves that the samples were discharged adequately after each round [15].

4 Results & Discussions

The presented results in this chapter are measured quantities. The represented graphs were plotted in a log scale based on the obtained measurement results. The mathematical representation of PDC has been given in Chapter 2. The polarization and depolarization currents given here are measured quantities by DIRANA. From these results using Equation 3, the change in the conductive current

due to ageing is obtained. Even though Equation 3 would allow the specific conductivity to be determined, considering the complexity of the cable structure, it would be difficult to calculate accurately [31]. Hence, the conductive current change with ageing was analyzed here. The conductive current has been calculated by subtracting measured polarization and depolarization currents. It is essential to mention that this paper discusses the initial stage of the ongoing research. As a result, no evaluation of performance has been conducted yet. To obtain a comprehensive performance measurement, additional ageing cycles would be required.

4.1 Core

Figures 7 and 8 illustrate the polarization and depolarization currents in a log scale over time at different ageing rounds. Figure 7 shows that the polarization current rose over ageing as the whole curve shows an increasing trend. It could be said that the polarization process has increased due to chemical and physical reactions [24]. It was already mentioned in the literature that the magnitude of PDC currents elevates with the temperature [32].

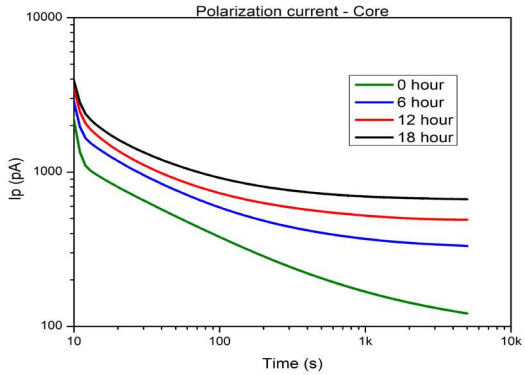


Figure 7
Polarization current – Core

The absolute values of depolarization current were used to plot Figure 8 for easier comparison. When the first and the last round is compared, i_d increased due to ageing up to 200 s. After this point, it goes even under the first round. The reason could be the diffusion of the plasticizer between PVC layers and new molecule bounds, which affects the conductivity [4], [6].

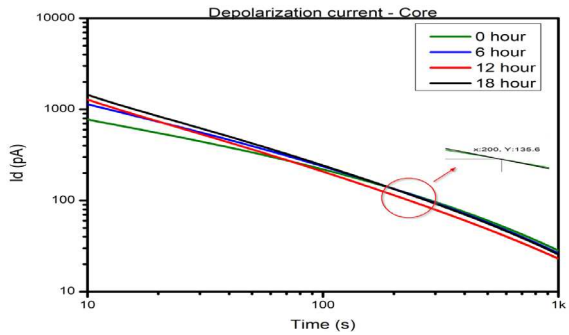


Figure 8
Depolarization current – Core

As given in Equation 3, the conduction current (i_c) can be calculated as the difference between the polarization and depolarisation currents. Based on this, the conduction current was calculated. The sampling time is taken as 1 second. I_c is calculated up to 1000 seconds, the duration of discharging period. Figure 9 depicts the conduction current versus time in the log scale for core insulation. It is visible that the conduction current has increased with ageing.

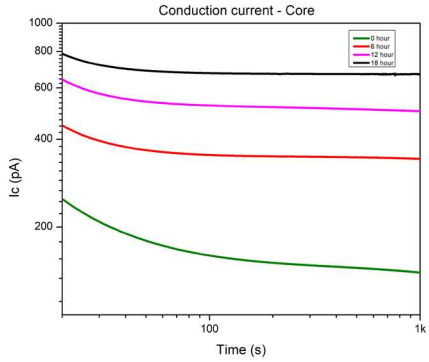


Figure 9
Conduction current – Core

4.2 Jacket

The curves of the PDC measurement of the jacket is shown in Figure 10 and 11. By looking at the results, it is possible to see a clear increasing trend for both i_p and i_d. The growing trend of polarization current indicates an increase in the polarization level [21]. It is also possible to observe the rise in the depolarization current. The reason could be that plasticizer evaporation from core insulation and PVC belt towards the jacket may raise the conductivity [4], [6], [7], [28]. As a result, the depolarization current has shown an increasing trend. However, this plasticizer behaviour would require further chemical studies.

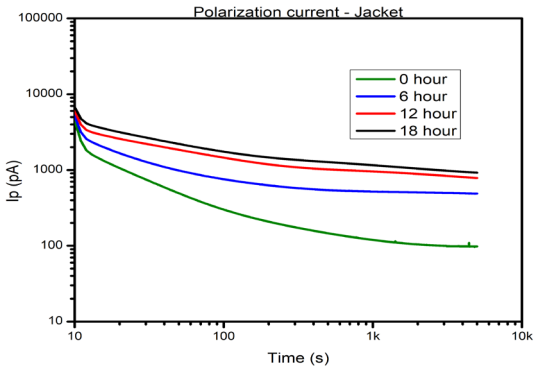


Figure 10
Polarization current – Jacket

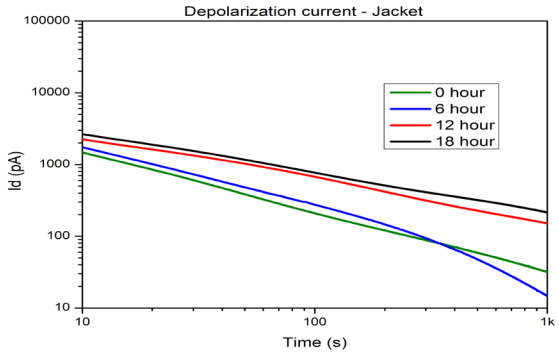


Figure 11
Depolarization current – Jacket

The conduction current of the jacket has also shown a similar trend to the core. It has increased after each ageing cycle. This trend can be seen in Figure 11. It is also noticeable from the chart that the conduction current has increased significantly after the first round of ageing compared to the unaged condition. However, the difference between the other rounds has shrunk with ageing. The same trend can be observed for the core, as well. As mentioned, the reason could be the diffusion of the plasticizer between the PVC layers.

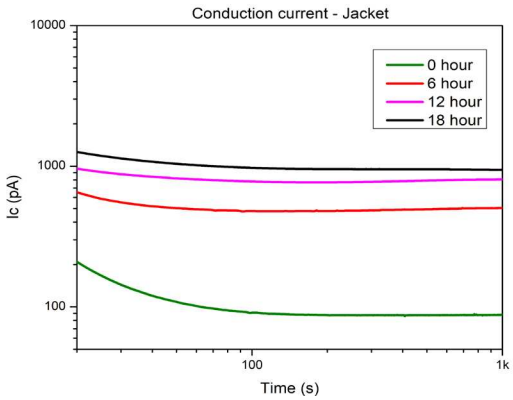


Figure 12
Conduction current – Jacket

Conclusions

The PVC insulated low voltage distribution cable samples were exposed to the short-term cyclic accelerated thermal ageing in the temperature-controlled oven at 110°C. Each round lasted six hours, and eighteen hours of total ageing time was reached. The equivalent ageing time of 330 hours was calculated using the Arrhenius equation. Non-destructive PDC measurement was used to explore the impact of thermal ageing on the cable insulation. The results so far show that ageing affects the polarization and depolarization current. The observed elevation in the polarization current for the core and jacket indicates the polarisation level's rise.

It is possible to see that the jacket's depolarisation current rises due to ageing. The migration of additives and plasticizers towards the jacket from core insulation and PVC belt might be the reason behind this. The depolarization current of the core has shown an unclear trend. It has increased due to ageing up to 200 s. After this point, it goes even under the first round. The reason could be that the new molecule bounds are being established, affecting conductivity. More ageing cycles are needed to get a clear understanding.

Acknowledgement

Project no. 142814 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the FK_22 funding scheme.

References

[1] B. Kruizinga, P. A. A. F. Wouters and E. F. Steennis, "Fault development on water ingress in damaged underground Low Voltage cables with plastic insulation," 2015 IEEE Electrical Insulation Conference (EIC), 2015, pp. 309-312

[2] K. Anandakumaran, S. Barreca, N. Seidl and P. V. Castaldo, "Nuclear qualification of PVC insulated cables," in IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 8, No. 5, pp. 818-825, Oct. 2001

[3] J. Packa, M. Váry, J. Lelák and V. Šály, "Degradation of 230/400 V distribution cables with PVC insulation," 2017 18th International Scientific Conference on Electric Power Engineering (EPE), 2017, pp. 1-4

[4] G. M. Csányi, S. Bal, and Z. Á. Tamus, "Dielectric Measurement Based Deducted Quantities to Track Repetitive, Short-Term Thermal Aging of Polyvinyl Chloride (PVC) Cable Insulation," Polymers, Vol. 12, No. 12, p. 2809, Nov. 2020

[5] "2030 Climate Target Plan." [Online] Available: https://climate.ec.europa.eu/eu-action/european-green-deal/2030-climate-target-plan_en

[6] S. Bal and Z. Á. Tamus, "Investigation of Effects of Thermal Ageing on Dielectric Properties of Low Voltage Cable Samples by using Dielectric Response Analyzer," 2022 International Conference on Diagnostics in Electrical Engineering (Diagnostika), 2022, pp. 1-6

[7] Csányi, G. M., Tamus, Z. Á., Kordás, P., "Effect of Enhancing Distribution Grid Resilience on Low Voltage Cable Ageing" In: Camarinha-Matos, L., Adu-Kankam, K., Julashokri, M. (eds) Technological Innovation for Resilient Systems. DoCEIS 2018, IFIP Advances in Information and Communication Technology, Vol. 521, Springer, Cham, 2018

[8] Son, N. N., Vinh, Luu The, "Parameter Estimation of Photovoltaic Model, Using Balancing Composite Motion Optimization", Acta Polytech. Hungarica, Vol. 19, No. 11, pp. 27-46

[9] Sami, T., A. Gholami, and S. M. Shahrtash. "Effect of electro-thermal, load cycling and thermal transients stresses on high voltage XLPE and EPR power cable life.", pp. 190-190, 2009

[10] V. Nikolic and R. Polansky, "Assessing the reliability of electrical insulating materials using accelerated aging tests," 2020 International Conference on Diagnostics in Electrical Engineering (Diagnostika), 2020, pp. 1-4

[11] S. Bal and Z. Ádám Tamus, "Investigation of Effects of Short-term Thermal Stress on PVC Insulated Low Voltage Distribution Cables", Period. Polytech. Elec. Eng. Comp. Sci., Vol. 65, No. 3, pp. 167-173, Jan. 2021

[12] A. Paramane, X. Chen, C. Dai, H. Guan, L. Yu, and Y. Tanaka, "Electrical insulation performance of cross-linked polyethylene/MgO nanocomposite material for ±320 kV high-voltage direct-current cables," Polym Compos, Vol. 41, No. 5, pp. 1936-1949, May 2020

[13] J. C. Fothergill, S. J. Dodd, L. A. Dissado, T. Liu and U. H. Nilsson, "The measurement of very low conductivity and dielectric loss in XLPE cables: a

possible method to detect degradation due to thermal aging," in IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 18, No. 5, pp. 1544-1553, October 2011

[14] Y. Zhang, X. Chen, H. Zhang, J. Liu, C. Zhang, and J. Jiao, "Analysis on the Temperature Field and the Ampacity of XLPE Submarine HV Cable Based on Electro-Thermal-Flow Multiphysics Coupling Simulation," Polymers, Vol. 12, No. 4, p. 952, Apr. 2020

[15] W. S. Zaengl, "Dielectric spectroscopy in time and frequency domain for HV power equipment. I. Theoretical considerations," in IEEE Electrical Insulation Magazine, Vol. 19, No. 5, pp. 5-19, Sept.-Oct. 2003

[16] N. Sunthrasakaran et al., "Application of polarization and depolarization current in high voltage insulator-A review," International Journal of Integrated Engineering, Vol. 12, No. 2, pp. 159-169, 2020

[17] D. Mishra, N. Haque, A. Baral and S. Chakravorti, "Assessment of interfacial charge accumulation in oil-paper interface in transformer insulation from polarization-depolarization current measurements," in IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 24, No. 3, pp. 1665-1673, June 2017

[18] S. Morsalin and B. T. Phung, "Dielectric response study of service-aged XLPE cable based on polarisation and depolarisation current method," in *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 27, No. 1, pp. 58-66, Feb. 2020

[19] A. A. Suleiman, N. Bashir, N. A. Muhamad, N. A. M. Jamail and Q. E. Kamarudin, "Polarization & depolarization current (PDC) measurement in cellulose insulating materials," 2013 Annual Report Conference on Electrical Insulation and Dielectric Phenomena, Chenzhen, China, 2013, pp. 1073-1076

[20] R. S. A. Afia, E. Mustafa, and Z. Á. Tamus, "Dielectric Spectroscopy of Low Voltage Nuclear Power Cables Under Simultaneous Thermal and Mechanical Stresses," Energy Reports, Vol. 6, pp. 662-667, Dec. 2020

[21] S. A. Bhumiwat, "Insulation condition assessment of transformer bushings by means of polarisation/depolarisation current analysis," Conference Record of the 2004 IEEE International Symposium on Electrical Insulation, 2004, pp. 500-503

[22] S. K. Ojha, P. Purkait, and S. Chakravorti, "Modeling of relaxation phenomena in transformer oil-paper insulation for understanding dielectric response measurements; Modeling of relaxation phenomena in transformer oil-paper insulation for understanding dielectric response measurements," IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 23, No. 5, Oct. 2016

[23] T. K. Saha and K. P. Mardira, "Modeling metal oxide surge arrester for the modern polarization based diagnostics," IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 12, No. 6, pp. 1249-1258, Dec. 2005

[24] R. S. A. Afia, E. Mustafa and Z. Á. Tamus, "Condition Assessment of XLPO Insulated Photovoltaic Cables Based on Polarisation/Depolarisation Current," 2020 International Conference on Diagnostics in Electrical Engineering (Diagnostika), 2020, pp. 1-4

[25] T. K. Saha and P. Purkait, "Investigation of polarization and depolarization current measurements for the assessment of oil-paper insulation of aged transformers," in *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 11, No. 1, pp. 144-154, Feb. 2004

[26] Küchler, A. and Bedel, T. (2001) Research note: Dielectric diagnosis of water content in transformer insulation systems. Euro. Trans. Electr. Power, 11: 65-68

[27] IEC 60502-1:2021 "Power cables with extruded insulation and their accessories for rated voltages from 1 kV (Um = 1,2 kV) up to 30 kV (Um = 36 kV) – Part 1: Cables for rated voltages of 1 kV (Um = 1,2 kV) and 3 kV (Um = 3,6 kV)" (2021)

[28] G. M. Csányi, Z. Á. Tamus, and Á. Varga, "Impact of distributed generation on the thermal ageing of low voltage distribution cables," IFIP Advances in Information and Communication Technology, 2017, Vol. 499, pp. 251-258

[29] S. Bal, A. G. N. A. Jasim and Z. A. Tamus, "Analyzing the Effect of thermal Ageing on Low Voltage Distribution Cables in Relation to their Structure," *2023 INSUCON - 14th International Electrical Insulation Conference (INSUCON)*, Birmingham, United Kingdom, 2023, pp. 209-214

[30] M. Ekelund, H. Edin, and U. W. Gedde, "Long-term performance of poly(vinyl chloride) cables. Part 1: Mechanical and electrical performances," *Polym Degrad Stab*, Vol. 92, No. 4, pp. 617-629, Apr. 2007

[31] S. Bal and Z. Á. Tamus, "Investigation of the Structural Dependence of the Cyclical Thermal Aging of Low-Voltage PVC-Insulated Cables," Symmetry, Vol. 15, No. 6, p. 1186, Jun. 2023

[32] I. Fofana, H. Hemmatjou and F. Meghnefi, "Effect of thermal transient on the polarization and depolarization current measurements," in IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 18, No. 2, pp. 513-520, April 2011

Safety of a Child, in a Vehicle Side Crash, with Three Restraint Systems

Abbas Talimian, Jan Vychytil, Luděk Hynčík, Tomasz Bońkowski

Biomechanical human body models, New Technologies Research Centre, University of West Bohemia, Univerzitní 2732/8, 301 00 Plzeň, Czech Republic, talimian@ntc.zcu.cz, jvychyti@ntc.zcu.cz, hyncik@ntc.zcu.cz, tomasz@ntc.zcu.cz

Abstract: Child safety is the main concern for parents, while driving a vehicle, with a child on board. Appropriate child restraint systems (CRSs) optimistically lessen injuries and reduce the mortality rate of children in accidents. This study compares child safety in a near-side crash with three different restraint systems: safety vest, and three- and four-point seatbelts. A six-year-old child is modelled by scaling a reference human model. The child is in an upright posture and located on a simplified child seat. The child seat itself is affixed to a deformable car seat that is made from fully deformable foam. The foam is validated to be used in crash analysis. The interior of the child seat is also covered by a layer of this foam. Seatbelts and the safety vest are made from standard seatbelt membrane material. Simulations were done for the near-side crash by a sled test regarding the side barrier crash pulse in the virtual performance solution (VPS). Outcomes indicate that a four-point seatbelt and safety vest, can provide relatively safer conditions for a child, in a near-side crash, than a three-point seatbelt, from a thorax injury point of view. This is the direct consequence of distributing the load over the thorax. Using a four-point seatbelt puts delay in the thorax injury and helps to use other safety means to reduce injuries on the body. The assumed geometry of the safety vest is not appropriate and causes the model's pelvis injury grade to be worse than that seen with three-point seatbelts.

Keywords: child model; Virthuman; child restraint systems; sled test; injury assessment

1 Introduction

Child occupant injury prevention in motor vehicles is one of the most important public health issues around the globe. Regardless of the continuous increase in the number of vehicles on the road, a downward trend is seen in children's death rate as a result of improvements in child restraint systems [4]. However, up to half of the traffic casualties in high-income countries, are children. For instance, in 2004 almost 40% of fatal road traffic deaths recorded in the Czech Republic were

related to children [20]. Consequently, designing suitable child restraint systems (CRSs) is a must in the developing car industry. Besides an applicable design, the correct usage of CRSs raises the chance of reducing fatal injury risk. Child safety seats could reduce the fatal injury risk of infants and toddlers in passenger vehicles by 71% and 54% respectively [29]. It is also important to select an applicable restraint system according to the child occupant's age and weight because it can effectively reduce the deaths of young children, even up to 70% [20].

Three restraint systems are widely used for children, in vehicles [25]. Rear-facing restraint is suggested for infants. It is quite common to have toddlers in forward-facing restraints. For older children, it is recommended to use a booster that is located on the vehicle seat [1]. The usage of a booster helps to fit the vehicle's seatbelt on the child's thorax. Of course, it is possible to use a combination of CRS types for some children as they grow up. Unfortunately, some vehicle users are not familiar with installing and using standard seatbelts with a child restraint system such as that for a toddler who is on a booster [22]. Incorrect usage of a seatbelt with a restraint system potentially increases the risk of injury for a child who is aged 11 years and below in a vehicle crash [19]. On the other hand, when CRSs are used correctly, it lowers children's death risk by 28%, in comparison with an occasion in which a child just fastens a vehicle seatbelt [5].

Children's body parts are in a state of growth and they are more vulnerable to injuries in vehicle crash scenarios, compared to adults [20]. Regarding the national highway traffic safety administration (NHTSA), children have to use boosters until they reach age 8 unless they are taller than 1.45 m [24]. According to [19], vehicle users are advised to put their children in front-facing CRS with booster seats to be able to use the seatbelts which are designed for adult passengers [2]. A seatbelt is an ideal one if it can securely distribute pressure over either a wider area of the child's chest or hard bones rather than soft internal organs during a crash [24, 28]. The pressure distribution over the child's chest efficiently decreases its body part injury in a crash. When a vehicle brakes, the interaction between a seatbelt and a child's posture will be problematic [3]. Therefore, two belt-positioning booster seats were tested to meet federal motor vehicle safety standards: the high-back and the backless or low-back booster seats. A belt-positioning booster seat is useful, as it positions a child in a vehicle seat in such a way that a standard belt will fit across the child's thorax and lap.

From this context, a safety vest conceptual design, is introduced based on current research [13]. The safety vest ends are connected to the child seat and have a zipper in front. It is easy for a driver to use. The driver can put a child on the seat and close the vest's front, with a zipper. In addition, the safety vest distributes pressure over a child's thorax while a child fastens a seatbelt –the shoulder strap applies extra pressure on some thorax segments which are under it. In the present study, a child model is located on a simplified child seat model. Subsequently, the

body's injury levels for the safety vest are compared with three-point and four-point seatbelts in a side crash.

The present study models a child's body, by scaling the reference model Virthuman [16]. This model is a useful and reliable model for having a fast calculation process and predicting the injury risk on different body parts in different crash scenarios [15]. It is a hybrid scalable virtual model and has a skeleton that is formed as a multi-body structure (MBS); therefore, it can be easily positioned in various seating postures such as upright, reclined, lying down, etc. The skeleton is connected to deformable segments which represent the model's outer skin [9, 30, 31]. One can model a wide spectrum of occupants depending on their gender, height, weight and age by using the Virthuman.

2 Method

The Virthuman model and seats which represent the vehicle interior are generated in the Visual Environment (VE) and calculated in the Virtual Performance Simulation (VPS) environment, the PAM-Crash module. Virthuman in this study models a 6-year-old male with 124.75 cm height and weight of 23 kg (53^{rd} percentile male) [16]. The body is located on a simplified child seat model. As far as the height of the child model is concerned, it is shorter than 145 cm, a child seat must be used for it [24] and is put on the vehicle seat. The child model is in an upright seating posture. The vehicle seat is also in an upright posture and its back angle with the global vertical axis (-z) in the VPS environment is 15.43 degrees.

For simulation simplicity reasons, the child seat is modelled by a shell frame made from ABS and has a 10 mm thickness. 40 mm polyurethane foam [26] covers the inner part of the frame. Its total weight is around 9.3 kg. The child seat was put on a vehicle seat which is also deformable [26]. The vehicle seat's geometry is based on the driver's seat of the Volvo XC70 Station Wagon [18] and its total weight is slightly more than 28 kg.

The foam is modelled using a general nonlinear strain-rate material in the VPS environment. It is a highly compressible non-linear elastic foam with strain-rate dependency and optional energy absorption (hysteresis). The foam behavior is described by two compressions and tension stress curves versus engineering strain, Figure 1. Its Young's modulus and unloading energy dissipation factors are 0.02 GPa and 0.25 respectively.

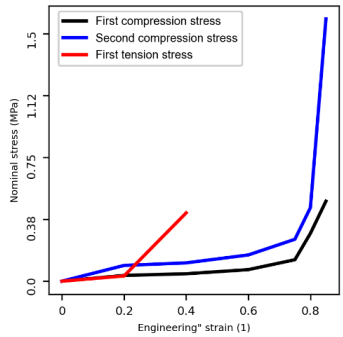


Figure 1
Polyurethane foam material

The child seat is connected to the vehicle seat with three elastic bars (a bar is on top of the child seat and two others are located on its back bottom) which have a 10 mm diameter. The inappropriate motion of the child seat, i.e., excessive bending of its back, is prevented by these bars.

Due to the usage of a deformable material in the present study for seats, a pre-simulation (free-fall) had to be done first to deform foam initially with the child's weight. Thus, the body model was released slightly above the seat from rest. The gravity acceleration is applied to all model nodes including the body, the child and the vehicle seats. By the end of pre-simulation nodes of the child model and seats are mapped to the initial model for representing the deformed models.

Instead of the seat's internal structure, the motion of chosen nodes under the seat cushion, back, headrest and armrest are fixed, Figure 2. Pure gravity acceleration was applied to the whole model along the z-axis for 500 ms as a pre-simulation, which was quite enough for the model to reach a stable position and deform the foam completely, as is shown in Figure 2.

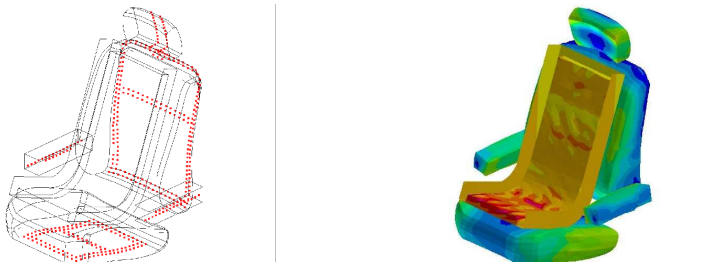


Figure 2
Vehicle seat. (Left) fixed nodes on the seat. (Right) deformed seat (pre-simulation).

Three child restraint systems (CRSs) that were selected for fastening the body to the seat are given in Figure 3.

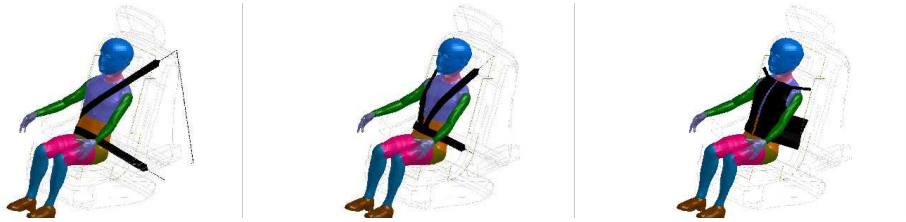


Figure 3
Different restraint systems. (Left) three-point seatbelt. (Middle) four-point. seatbelt (Right) safety vest.

The first case is an integrated standard continuous three-point belt. It has three key components: connecting bars, and shoulder and lap membranes. This type of seatbelt is connected to a vehicle structure at three points: D-ring, Buckle and Anchor. The shoulder belt is started from the D-ring and passes over the trunk diagonally from upper left to lower right (from D-ring to Buckle). Meanwhile, the lap belt is located over the belly from right to left (from Buckle to Anchor). Non-linear bars are used to connect the free ends of membranes to external parts of a seatbelt (Retractor, D-ring, Buckle and Anchor). The nonlinear bar segment starts from the retractor, passes through the D-ring and ends at the upper part of the shoulder membrane. Another nonlinear bar part connects the shoulder and lap membranes via the Buckle. Finally, the bar segments connect the other side of the lap belt to the Anchor. Restraint systems that are considered for the present study are made from standard seatbelt materials, Figure 4. The membrane belt's width is 40 mm and its thickness is 1.2 mm.

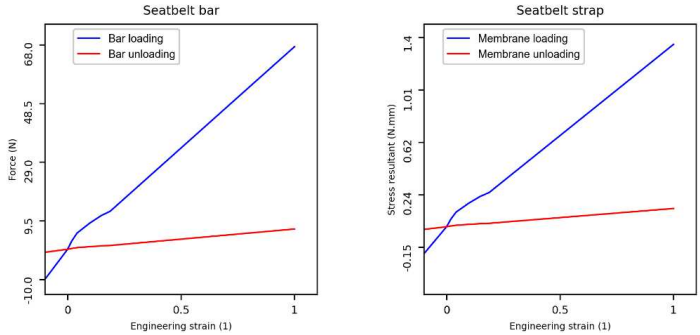


Figure 4
Seatbelt material. (Left) Seatbelt bar. (Right) Seatbelt strap.

A four-point seatbelt is the second restraint system which was considered in the present study for fastening the child model to the child seat. In the four-point seatbelt, membranes are attached to the child seat frame at four points by nonlinear bars rather than to the vehicle structure. The membrane material and bar material are the same as what is used for the three-point seatbelt. The safe tools in

the Virtual Performance Simulation (VPS) environment were used for generating three- and four-point seatbelts.

The third restraint system chosen is a conceptual design that is called a safety vest. The vest model has two panels which are attached to the child seat sides. The safety vest's panels are connected right in front of the model chest by a zipper. The safety vest is made from standard seatbelt material with 1.2 mm thickness the same as the three- and four-point seatbelts. The vest zipper is modelled by bars whose material properties are identical to the nonlinear bars used in the three-point seatbelt. The safety vest was fitted to the model skin's outer surface. The minimum initial vest offset is approximately 1mm between the thorax's side part and the safety vest panels, while the maximum initial offset is 9 mm and is between the thorax's frontal part and the safety vest's panels.

3 Crash Simulation

Since 2016, in accordance with the Euro NCAP regulations, larger child occupant dummies have to be considered for side crash tests [8]. The side crash test can be done by applying a crash pulse either to the model [21] or the impact barrier [27]. For evaluating passive safety system performance, i.e. seatbelts and airbags, it is common to use sled tests [10]. Either a full car body or a part of it, depending on the research goal, is located on a rail and the same deceleration pulse is applied to the model. Whenever occupants are situated in lightweight vehicles, side crash scenarios are a challenging issue, because they are faced with high-value pulses [27]. The Euro NCAP side-impact barrier crash acceleration pulse is used for simulating a side crash, Figure 5 [12, 23].

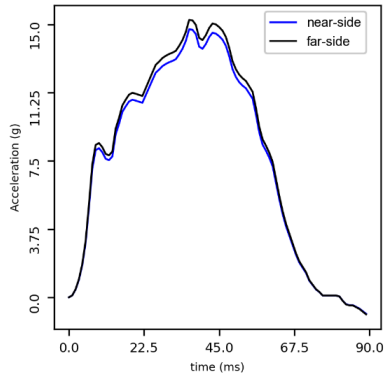


Figure 5
Acceleration pulse

For simulating a near-side crash, the acceleration pulse is applied horizontally to the model's center node along the global y-axis in the VPS. This center node is in connection with fixed nodes, Figure 2, as multiple nodes to one node constraint (MTOCO). The model's center node is just free to move along the y-axis (the crash axis); meanwhile, its other degrees of freedom are fixed. The three-point seatbelt parts are not symmetric like the two other restraint systems (four-point seatbelt and safety vest). Therefore, a far-side crash test must also be done. For this test, the acceleration pulse is applied horizontally to the model's center node along the global y-axis but in the negative direction.

4 Body Part Injury Assessment

A program is available for the Virthuman model to do post-processing and assess the particular model part (head, neck, thorax, abdomen, pelvis, femurs, knees and tibiae) injury level [15]. The injury level of each part is computed according to its kinematics and kinetics during the crash. Accelerations, forces and displacements are measured for joints that connect Virthuman's outer surface to its MBS structure. For instance, the head's center of gravity is selected to store the acceleration component and subsequently, the head injury criterion (HIC) is computed based on the absolute value of the head's acceleration. A model's neck injury level is determined by measuring three factors: flexion/extension moment, tension/compression and shear force of the neck joint. Deflections of segments on the model's thorax are measured during the simulation and compared with the existing injury criteria database for evaluating the thorax's injury level. The abdomen and pelvis injury levels are found based on compression and pubic forces respectively. Compression force and moment are collected for femurs and tibiae parts to determine these parts' injury levels. The injury level of knee parts is measured according to their joint's moment.

From the literature, injury criteria metrics are available for Virthuman's body parts (a 6-year-old child, a 20-year-old and a 100-year-old adult) [7, 11, 17]. It has already been mentioned that Virthuman is a scalable model. Therefore, the injury criteria are interpolated linearly according to the current passenger age at first by the post-processing program. Later on, either kinematic or kinetic data that were collected for body parts as output time-dependent variables are compared to the modified injury criteria for each body part. A body part injury level for a specific time is the worst injury level of a body part in the time interval from the beginning of the simulation to that time [14].

The overall injury level is graded at four basic levels (color-coded) in the software viewer in a similar way to the Euro NCAP consumer rating [6]. Either a small degree of injury or none is given by "Good". An injury level can also be "Acceptable" or "Marginal". However, injuries are shown by the "Poor", which

represents a very serious degree of injury, Figure 6. For instance, a head injury criterion is determined based on the HIC36 value. If it is lower than 650, a “Good” level is assigned to the head and it is represented by a “Green” color on the model in the VPS visual environment (VE). In contrast, in a case where HIC36 is more than 1000, the program returns “Poor” and colors the head in VE with “Red”. All body injury criteria are available in [15].

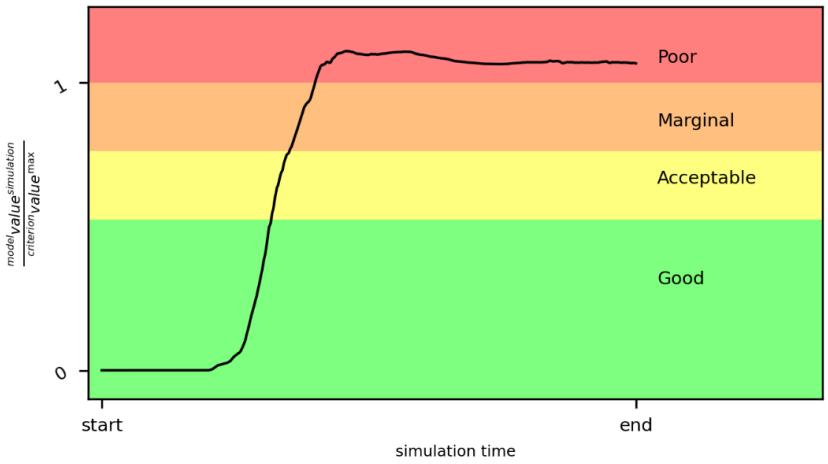


Figure 6
Injury assessment for a body part

5 Discussion

The kinematics of the child occupant body for the side crash tests (near- and far-side) are illustrated in Figure 7, regarding the three restraint systems.

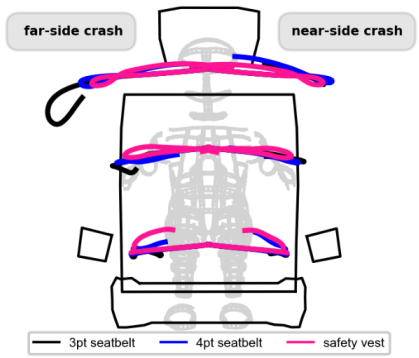
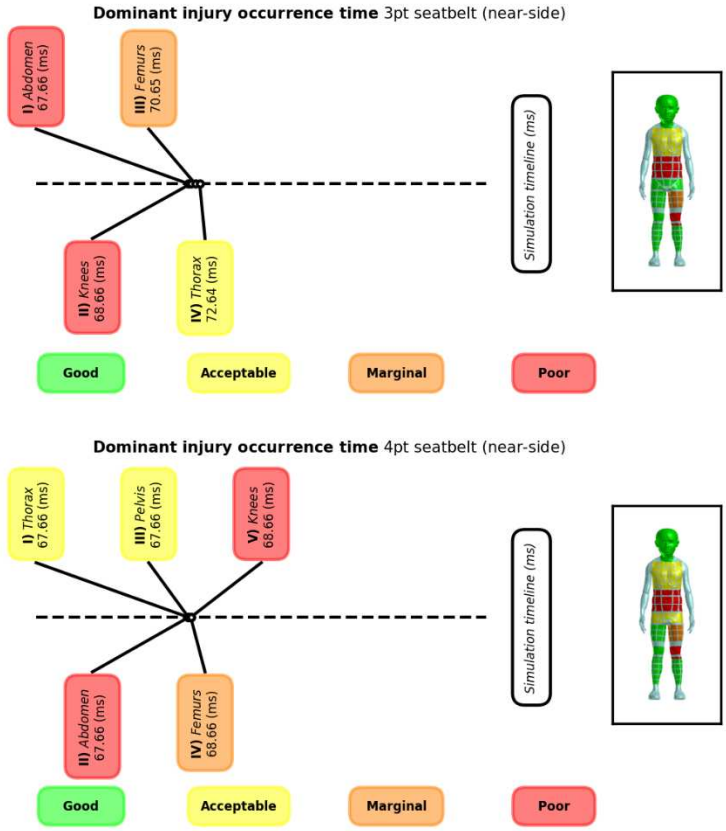


Figure 7
Child occupant kinematics in side collisions with different restraint systems

Curves in plots show the trajectory of nodes on the head center, sternum and hip in crashes. The curves which are located on the right side of the figure illustrate the trajectory of mentioned nodes in the near-side crash. The nodal trajectory on the left side of Figure 7 relates to these nodes in the far-side crash. For restraint systems that have symmetric geometry (four-point seatbelt and safety vest) Figure 3. The body's kinematic response is also symmetric in near- and far-side crashes Figure 7. The three-point seatbelt does not have symmetric geometry; hence, the body's kinematic responses are not symmetric in side crash scenarios Figure 7. Therefore, from now on body injuries are discussed for restraint systems in the near-side crash and far-side crashes with a three-point seatbelt.

The occurrence time of the final injury level – the dominant one – per body part is illustrated in Figure 8 for each restraint system. The simulation timeline is represented with a horizontal dashed line. Body parts with a "Good" injury level are not shown in occurrence time plots and the plot just shows injury occurrence times for the parts which have other level injuries, i.e., "Acceptable", "Marginal" and "Poor".



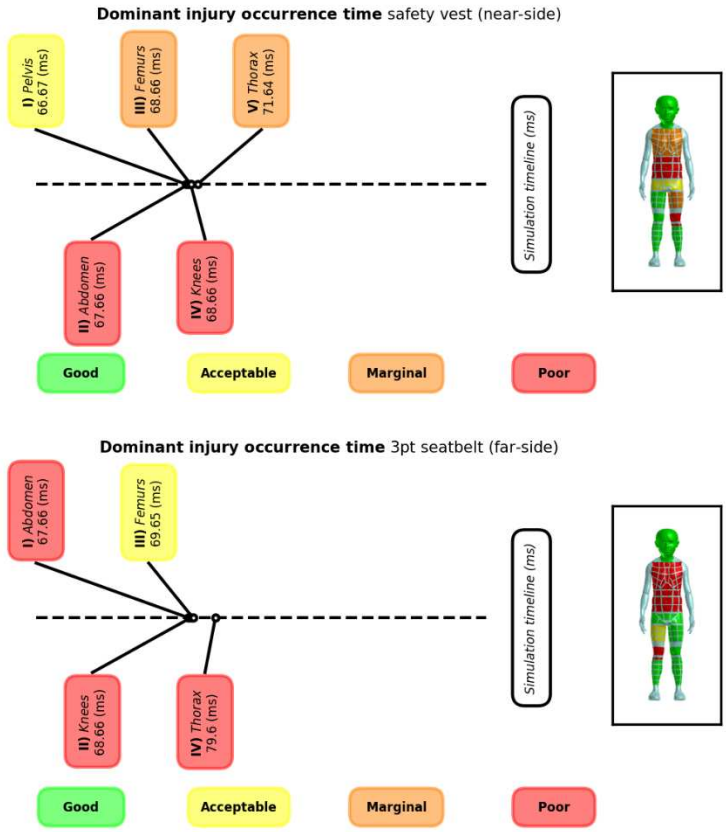


Figure 8

Dominant injury criteria occurrence time for a model in the side crashes

A detailed vehicle interior is not considered in this study and the body is subjected to side crash pulses; hence, the model head moves laterally and does not collide anywhere. In consequence, there is no sudden jump in the head acceleration curve versus time. Therefore, the head injury criterion (HIC36) is 44.89, 45.34 and 59.1 for the three-point seatbelt, four-point seatbelt and safety vest in the near-side crash, respectively. HIC36 is 52.7 for the model with a three-point seatbelt in the far-side crash Figure 9. All of these values are lower than 650 [15]; hence, the head injury level for these test occasions is “Good” and is indicated by “green” in the software viewer Figure 8.

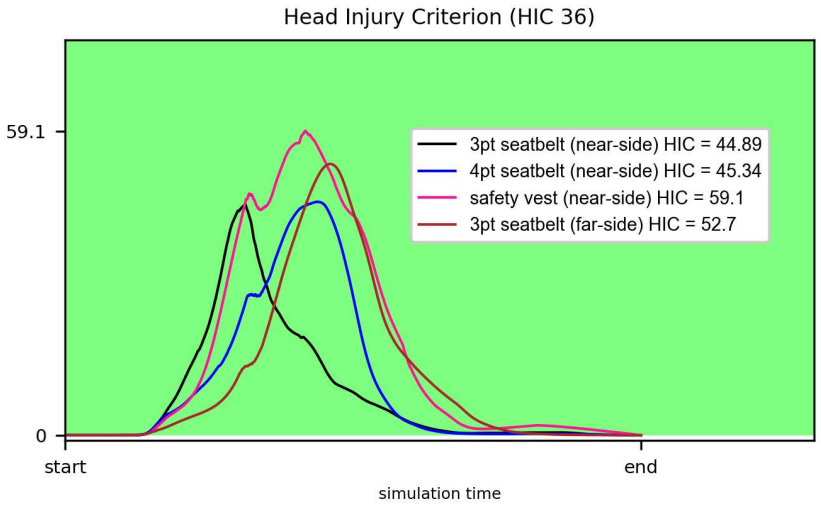


Figure 9
Non-dimensional neck extension moment in relation to simulation time

Information can be found in the colored boxes pointing to the timeline with a dark line. The injury sequence for a part is given in roman numerals. The first moment a part's dominant injury started is also given in the boxes. As an example, in Figure 8, with the three-point seatbelt in the near-side crash, the abdomen and knees reach "Poor" injury levels at 67.66 and 68.66 ms, respectively. Its femur injury level will be "Marginal" since it is 70.65 ms while the thorax reaches an "Acceptable" injury level of 72.64 ms. For ease of viewing, rounded numbers are put in the boxes. From the dominant injury occurrence time point of view, the majority of the body parts face a level of injury at around 70 ms with a side crash. The simulation's time, as well as the duration of the crash pulse's time, is 200 ms. Comparing dominant injuries in near-side crashes indicates that the most severe of injuries involve the knees and abdomen, with "Poor" levels and occur around 68 ms.

For side crashes, the model is subjected to lateral motion and bends to its side as a consequence of the inertia distribution. Restraint systems restrict the abdominal and thoracic motions but the difference between the model's trunk and head inertia generates a moment on the model's neck. The neck moment in the sagittal plane concerning simulation time is illustrated in Figure 10. None of the moment curve related to restraint systems exceeds the "green" band and the post-processing algorithm detects the "Good" injury level for the neck that can be seen in Figure 10.

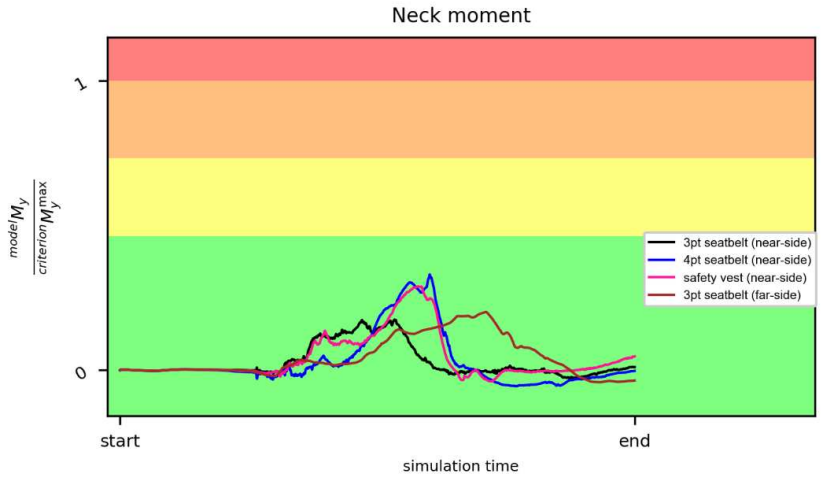
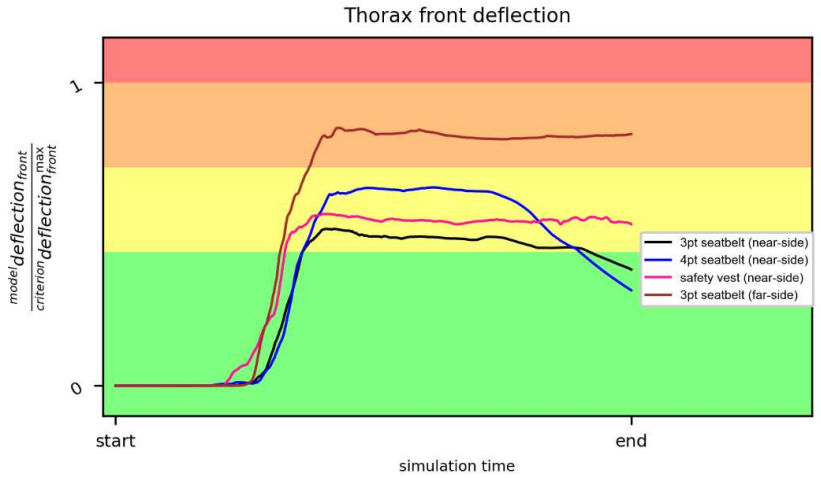


Figure 10
Non-dimensional neck extension moment in relation to simulation time

The post-processing takes the thorax deflections at frontal and side segments into account to determine the injury level of the thorax. Variations of the thorax deflection vis-a-vis simulation time are given in Figure 11. There are several segments on the Virthuman thorax. A thorax segment, which has the highest deflection peak is illustrated in Figure 11, as it is the most effective part for determining the thorax injury level. From the plots, it is seen that a model that has a three-point seatbelt, has the worst injury level, “Poor”, in the far-side crash Figure 8.



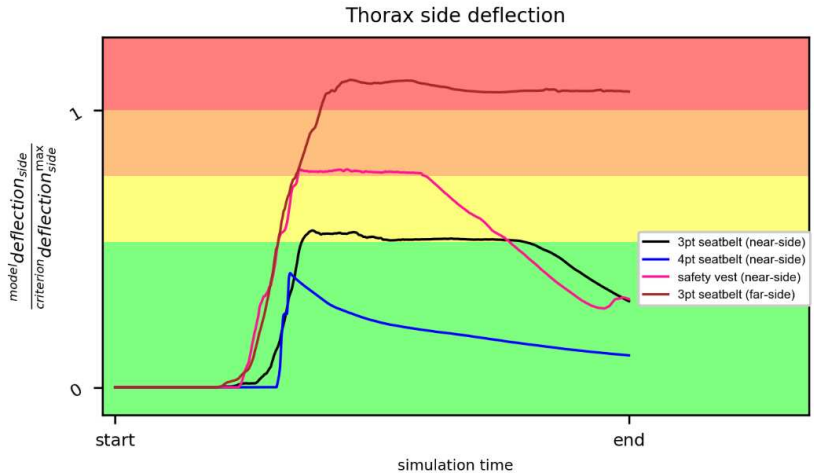


Figure 11
Non-dimensional thorax deflection in relation to simulation time. (Up) frontal segments. (Down) side segments.

This is the result of a large deflection on the thorax side segment Figure 11. The model with a safety vest in the near-side crash has a "Marginal" injury level because of the thorax deflection that is in the "brown" band. For the rest of the test cases, the "Acceptable" injury level is assigned to the thorax because a part of the thorax response curve is located in the "yellow" band.

Variation of the compression force for an abdominal segment in relation to simulation time is plotted in Figure 12. This segment has the highest compression force peak value among the other abdominal segments. Hence, its effect in determining the abdomen's injury is more than with other segments. Comparing the abdominal non-dimensional force regarding different restraint systems indicates that none of these restraint systems can provide a safe condition for the model from an abdominal injury point of view. Hence, in Figure 8 the abdomen is shown in "red".

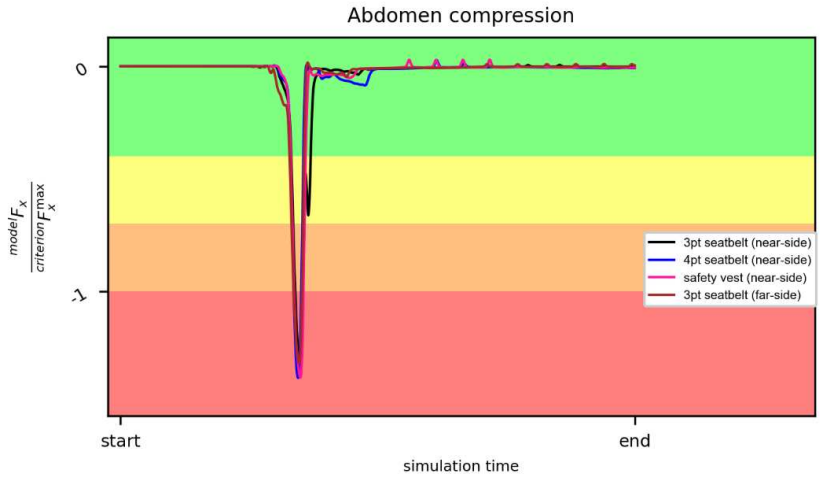


Figure 12
Non-dimensional abdomen compression in relation to simulation time

The pubic force in relation to simulation time is demonstrated in Figure 13 for each restraint system. The three-point seatbelt in the far-side crash has the best reaction compared to the others. When the near-side crash is the test's concern, the three-point seatbelt reaction is acceptable and keeps the pubic force peak in the "Good" injury level domain. However, the pubic force peak for the four-point seatbelt and safety vest is slightly located in the "yellow" band and the post-processing program indicates an "Acceptable" injury level for the model pelvis in these scenarios.

In Figure 14, one has the variation of the femoral moment in relation to simulation time. All of the restraint systems act quite similarly in the near-side crash and the femur injury level is determined as "Marginal" by the post-processing algorithm. The three-point seatbelt shows a better reaction in far-side crashes, as the femoral injury level is "Acceptable", which is better than the others in a near-side crash.

The knee joint moment variation in relation to time is illustrated in Figure 15 for different restraint systems. It is seen that the curves' peaks are in the "Red" zone and the knees' injury level for all restraint systems is "Poor". The type of restraint system has no considerable effect on knee injuries because it cannot restrict the motion of the body's lower extremities.

The post-processing algorithm deals with tibiae moment in relation to simulation time to evaluate their injury level. According to Figure 16, when either model is subjected to a near- or far-side crash, its tibiae injury level will be "Good" and restraint system types act in a similar way.

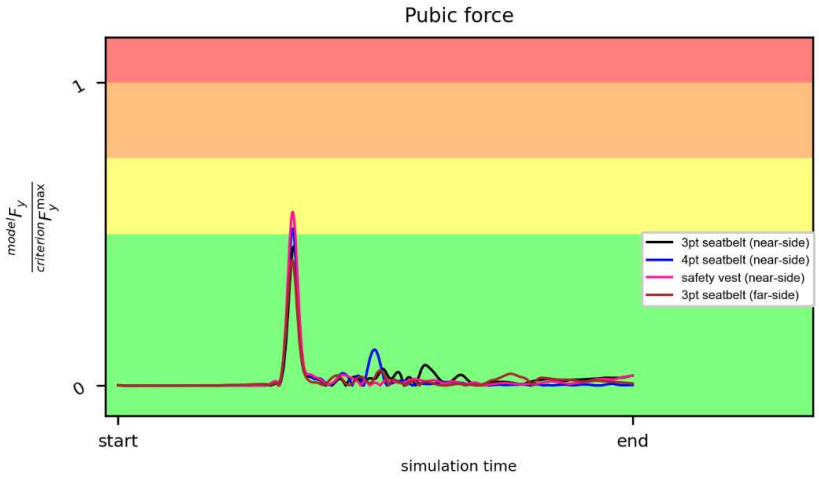


Figure 13
Non-dimensional pubic force in relation to simulation time

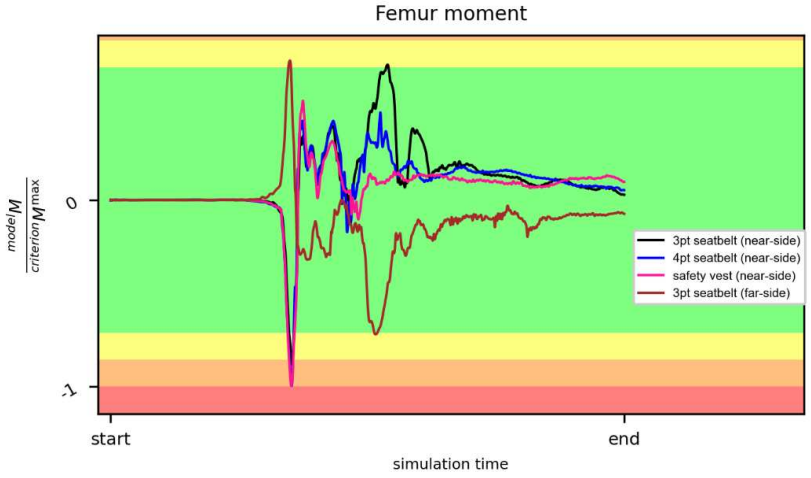


Figure 14
Non-dimensional femoral moment in relation to simulation time

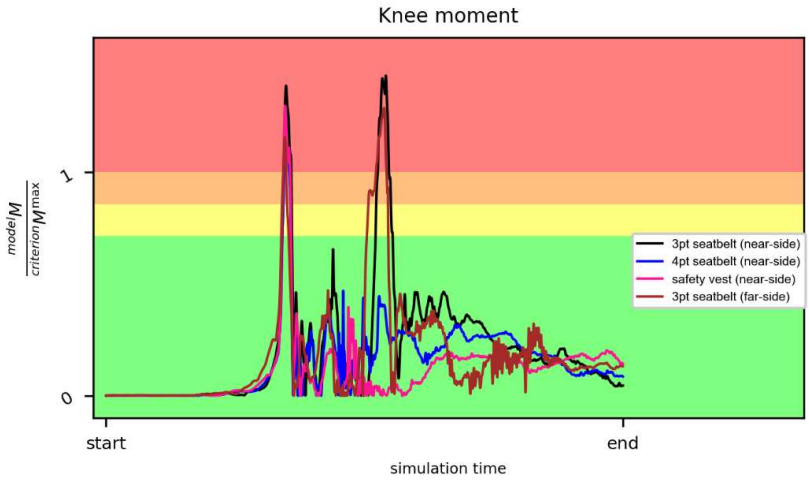


Figure 15
Non-dimensional knee moment in relation to simulation time

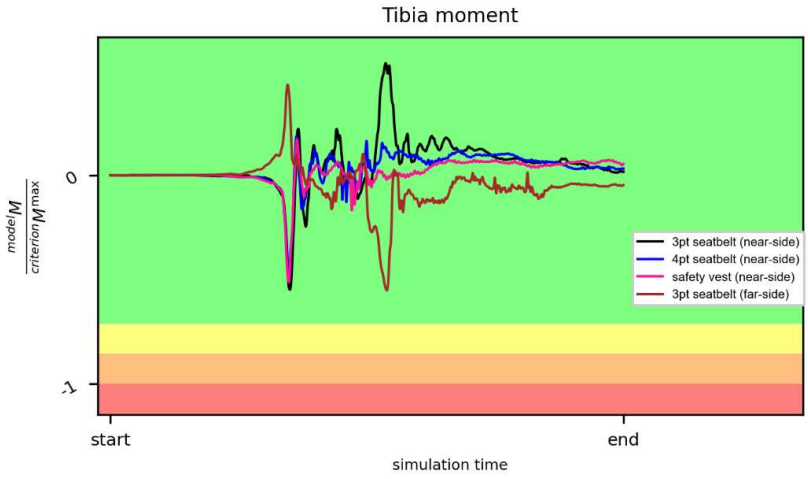


Figure 16
Non-dimensional tibia moment in relation to simulation time

The safety restraint systems in the present study are made from standard seatbelt materials, Figure 4, and their thickness is 1.2 mm. The sectional force of a belt or safety vest is measured for a section on the top left of the strap. Figure 17 shows the position of the section in red for the three-point seatbelt as an example of where strap force is recorded during the crash simulation. The variation of section forces in relation to the time is given in Figure 18 for different restraint systems. Because the force is distributed over a wider area (the contact area between the restraint system and the child's thorax) it causes lower force applies to the four-

point seatbelt and safety vest strap, that attaches the vest to the seat structure, in comparison with the three-point seatbelt. It can be seen that the section forces' peak for three- and four-point seatbelts are almost three times higher than is measured for a safety vest.

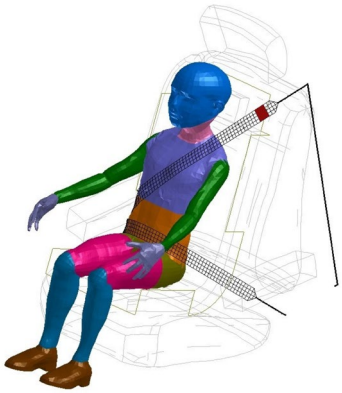


Figure 17
Strap section force position

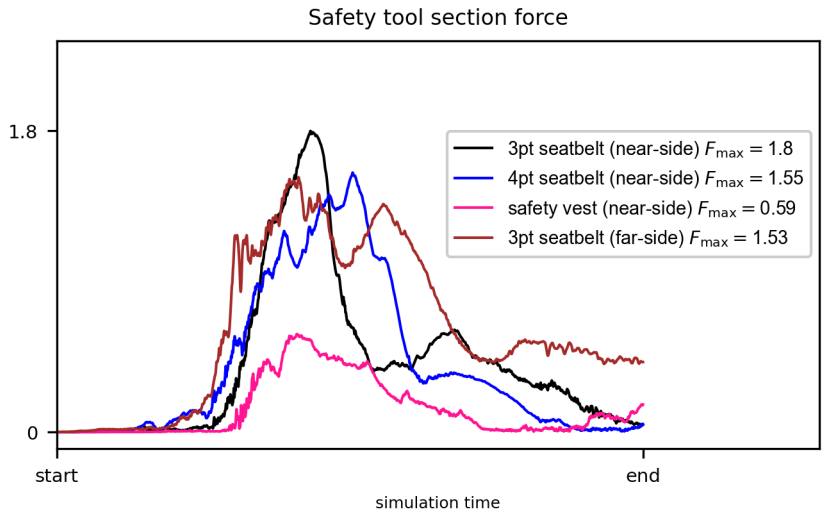


Figure 18
Strap section force

Seatbelts (three- or four-point seatbelts) and safety vest are not the only components which restrict a child's model motion. The child model is on a seat made from polyurethane foam that absorbs a part of crash energy. When the child model has either a four-point seatbelt or safety vest, the motion of the child's thorax is restricted more effectively than on occasion it has a three-point seatbelt.

The amount of crash energy that is dissipated by the seat's foam in these cases is relatively lower than the test case done by a three-point seatbelt. It happens because a child is fastened by the four-point seatbelt and safety vest its trunk has less lateral rotation than a case it has a three-point seatbelt.

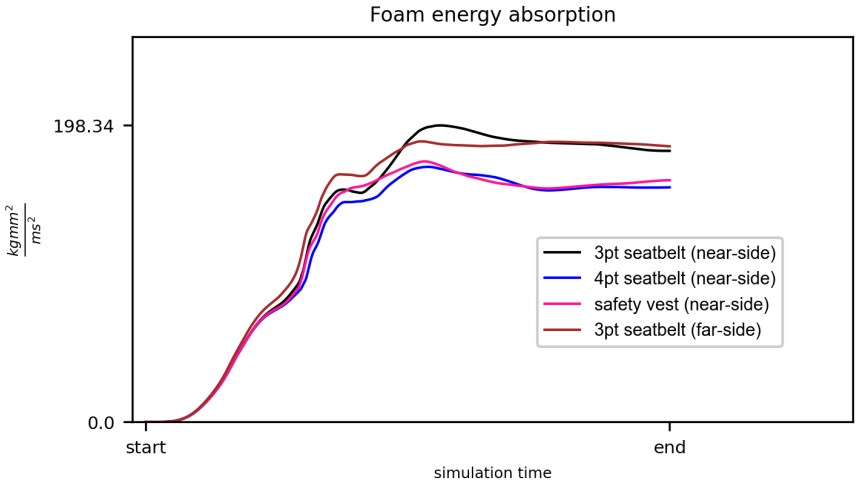


Figure 19
Seat energy absorption

Conclusions

A comparative study was completed to consider the influence of using three restraint systems, for a child occupant, in a side crash. A 6-year-old (53rd percentile male) reference Virthuman model is scaled to represent the child as a model. The body was in a simplified child seat, that is positioned above a deformable vehicle seat. Seats are made from polyurethane foam. The simulation was run in the Virtual Performance Simulation (VPS) environment first and an algorithm which is available for Virthuman processed the results to determine the injury level for particular body parts. The symmetrical geometry of the four-point seatbelt and safety vest causes the kinematics of the body in near- and far-side crashes to be symmetric as well. But the three-point seatbelt's geometry is not symmetric like the other restraint systems. Therefore, it is expected not to have symmetric kinematic curves, for a body in near- and far-side crashes. From the particular injury level point of view, a four-point seatbelt and safety vest cannot provide safer conditions for a child model encountering a side crash, in comparison with a model, which has a three-point seatbelt. However, by using a four-point seatbelt, a delay occurs in the thorax injury, which helps other safety means to reduce injuries to the body. The safety vest geometry is shown, in this study, not to have an appropriate design. Therefore, it allows serious pelvis injury to occur and is not as good as the three-point seatbelt.

Acknowledgement

This work was supported by the European Regional Development Fund-Project "Application of Modern Technologies in Medicine and Industry" (No. CZ.02.1.01/0.0/0.0/17_048/0007280).

References

[1] *Car Seat Recommendations for Children*. Available from: https://www.nhtsa.gov/equipment/car-seats-and-booster-seats.

[2] Arbogast, K. B., et al., *Child occupant protection: a summary of current safety recommendations.* Primary care update for Ob/Gyns, 2001. **8**(4): p. 141-148.

[3] Baker, G., et al. *Kinematics and shoulder belt engagement of children on belt-positioning boosters during emergency braking events*. in *The International Conference on the Biomechanics of Impact (IRCOBI)*. 2017.

[4] Ebel, B. E. and D. C. Grossman, *Crash proof kids? An overview of current motor vehicle child occupant safety strategies.* Current problems in pediatric adolescent health care, 2003. **33**(2): p. 38-55.

[5] Elliott, M. R., et al., *Effectiveness of child safety seats vs seat belts in reducing risk for death in children in passenger vehicle crashes.* Archives of pediatrics adolescent medicine, 2006. **160**(6): p. 617-621.

[6] EuroNCAP, ASSESSMENT PROTOCOL – ADULT OCCUPANT PROTECTION, 9.1.2, 2020.

[7] Forman, J. L., et al. *Fracture tolerance related to skeletal development and aging throughout life: 3-point bending of human femurs*. in *The International Conference on the Biomechanics of Impact (IRCOBI)*. 2012.

[8] Holtz, J., et al. *Side-impact simulation study to investigate the protection of older child occupants in lightweight vehicles*. in *The International Conference on the Biomechanics of Impact (IRCOBI)*. 2016.

[9] Hyncik, L., et al., On scaling virtual human models, 2013.

[10] Jimenez, M., et al., *Finite element simulation of mechanical bump shock absorber for sled tests.* International Journal of Automotive Technology, 2015. **16**(1): p. 167-172.

[11] Kleinberger, M., et al., *Development of improved injury criteria for the assessment of advanced automotive restraint systems.* 1998. **4405**(9): p. 12-17.

[12] Léost, Y. and M. Boljen. *Crash simulations of electric cars in the EVERSAFE project*. in *COMPLAS XIII: proceedings of the XIII International Conference on Computational Plasticity: fundamentals and applications*. 2015. CIMNE.

[13] Luttenberger, P., et al., *Assessment of future occupant restraint principles in autonomous vehicles.* The International Conference on the Biomechanics of Impact (IRCOBI), 2020.

[14] Mecas, Virthuman Postprocessing Manual (VPS Explicit MBS Model, Model Version 1.1) Rev. 1, 2016.

[15] Mecas, Virthuman Postprocessing Manual (VPS Explicit MBS Model, Model Version 1.5) Rev. 1, 2020.

[16] Mecas, Virthuman User's Manual (VPS Explicit MBS Model, Model Version 1.5) Rev. 1, 2020.

[17] Mertz, H. J. and L. M. Patrick, *Strength and response of the human neck.* 1971: p. 2903-2928.

[18] Ozmumcu, S. *Volvo xc7 Station Wagon Seat*. 2012 [cited 7-8-2012]; Available from: https://grabcad.com/library/automobile-seats.

[19] Paiman, N. F., et al., *A study on the use and misuse of child restraint system (CRS) in Malaysia.* Journal of the Society of Automotive Engineers Malaysia, 2018. **2**(1): p. 5-13.

[20] Peden, M. P., et al., *World report on child injury prevention*. 2008, Geneva, Switzerland.

[21] Pintar, F. A., et al. *Worldsid assessment of far side impact countermeasures*. in *Annual Proceedings/Association for the Advancement of Automotive Medicine*. 2006. Association for the Advancement of Automotive Medicine.

[22] Posuniak, P., et al. *Child restraint systems: problems related to the safety of children transported in booster seats (without integral safety belts)*. in *2018 XI International Science-Technical Conference Automotive Safety*. 2018. IEEE.

[23] PROGRAMME, E.N.C.A., FAR SIDE OCCUPANT TEST & ASSESSMENT PROCEDURE, 2022.

[24] Reed, M. P., et al., *Effects of vehicle seat and belt geometry on belt fit for children with and without belt positioning booster seats.* Accident Analysis and Prevention, 2013. **50**: p. 512-522.

[25] Schroeder, P. A., *The Effects of Child Restraint System Use and Motor Vehicle Collision Severity on Injury Patterns and Severity in Children 8 Years Old and Younger*. 2018.

[26] Špirk, S., M. Křížek, and Š. Jeníček. *Polyurethane foam behaviour during impact*. in *MATEC Web of Conferences*. 2018. EDP Sciences.

[27] Stein, M., et al. *Concept for lateral Impact Protection of a centred Driver in a light Electrical Vehicle*. in *The International Conference on the Biomechanics of Impact (IRCOBI)*. 2014.

[28] Tarriere, C. *Children are not miniature adults*. in *Proc. of Int. Conf. on the Biomechanics of Impacts (IRCOBI)*. 1995.

[29] Transportation, U.S.D.o., National Highway Traffic Safety Administration (NHTSA) (2001) Traffic Safety Facts 2001. Children (DOT HS 809 471)

[30] Vychytil, J., et al., Prediction of injury risk in pedestrian accidents using virtual human model VIRTHUMAN: real case and parametric study, 2016.

[31] Vychytil, J., et al., Scalable multi-purpose virtual human model for future safety assessment, 2014.

Product Development: Imagines versus Practice

László Berényi[1], László Soltész[2]

[1]Institute of Management Science, Faculty of Economics, University of Miskolc, Miskolc-Egyetemváros, 3515 Miskolc, Hungary, laszlo.berenyi@uni-miskolc.hu

[2]Institute of Machine and Product Design, Faculty of Mechanical Engineering and Informatics, University of Miskolc, Miskolc-Egyetemváros, 3515 Miskolc, Hungary, laszlo.soltesz@emerson.com

Abstract: The increasingly agile environment requires quick responses, including technological solutions and project management decisions. Former investigations aimed to explore expert opinions about project success factors. This study enhances the investigations to student opinions for establishing a more comprehensive view of the critical issues. A harmonious collaboration based on agreed values and methods has been appreciated. The study used a survey among engineering and business students (n=155) about the success factors of product development projects. The goal was to explore their opinions and compare those with expert opinions. The results show that clear goals are considered very important by both groups. Project meetings, cooperation, and other soft factors are recognized as more important by the experts than by the students. The results confirm that the order of the opinions is significantly different. The cluster analysis marked three distinctive patterns of opinions. Improving the performance of the mixed project teams reflected in project success calls for action. Accepting the more professional approach the experienced project team members, the lack of students' competencies can be designated. Changes must be supported by training activities and rethinking the curricula in the field.

Keywords: new product development; project success; student opinion; IPA analysis

1 Introduction

The Oslo Manual 2018 [1] defined product and business process innovations by the object. These two objects must be interrelated with each other in practice. New product development and performance improvement of current products (including services) can be considered the initial point of finding diffusion opportunities and adjusting business processes. It is to note that from a businessman's viewpoint, the process changes may provide input for the directions of product development. A special process is the product development process since both product and organizational-level benefits become available by rethinking it. Product development literature agrees that challenges are enhanced, and the changes have

accelerated in recent years. The accelerated changes lead to an unpredictable environment, quickly changing product requirements. Consequently, a shorter lead time for the new product. In a different interpretation, the intensity, extension, and predictability of the environment force a change in strategy [2], including the rethinking of product development. The PMI [3] describes the phenomenon of uncertainty that can appear as the uncertainty of the requirements and uncertainty of technical degree (Figure 1) and offers a map to select the appropriate selection of managerial tools.

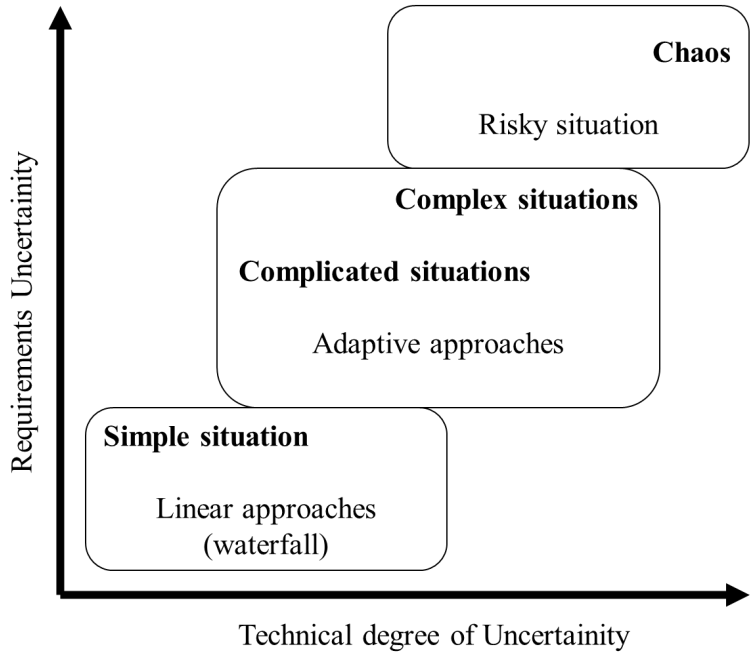


Figure 1
Applicable project management approaches by uncertainty (based on [3])

Baccarini [4] highlights the product success covering the organizational expectations as a success factor beyond the project management success. There is a change of focus point over time. The focus of attention has moved from project management success (in the 1960s–1980s) to project/product success (in the 1980s–2000s), then to project/product, portfolio, and program success and narratives of success and failure in the 21st Century [5].

Product development is usually managed by projects. Since its success goes beyond the project's outcomes, the organization's profitability or survival is in question; the classic iron triangle of project success (keeping time, keeping budget, delivering as planned) is not sufficient for understanding the success. Pollack et al. [6] show that the interpretation and the content of the triangle have been changing over time. They conclude that project success can be derived from the client's satisfaction with the

outcomes and impacts. Contributing to corporate strategy and stakeholders' satisfaction must be considered [7] [8]. Validation of many aspects derived from corporate and stakeholder-level expectations is a restraint of project management. In the case of product development projects, the forced lead time reduction is decisive, which requires a renewed approach to the process [9]. Regardless of the industry and the purpose of the project or the composition of the project team, project management remains the application of knowledge, skills, and techniques to project activities to meet project requirements [3]. The project manager is the person who is responsible for coordinating the related activities in all circumstances.

The new conditions affect the related project management approaches and methods [10]. The strengthening of design thinking methods [11] [12] [13] supports the appropriate responses of the product development teams to the challenges. Agility in project management seems to become general, even beyond software engineering [3] [14] [15] [16], and the values and methods appear in the product development approach as well. Product development models shifted gradually from the traditional predictive to an iterative or incremental project approach; however, the development is ongoing.

The emergence of new information technologies and their impact on the business model is admitted by the Oslo Manual [1]. The spread of information technologies also forces the development of new products and business models. Ultimately, these technologies determine the new models of product development.

Over and above the technical issues, one more enabler must be considered for the success of product development: competent human resources. The reduced lead time and the shorter product life cycles lead to an intensified workload on the development teams. Corporate capacities must follow the expansion both in quantity and quality. Involving recent professionals and higher education students is fundamental for assuring the recruitment and selection of future staff. Moreover, novel ideas and improved dynamics can be presented in the projects. Nevertheless, this situation may amplify conflicts and require increased project management coordination efforts.

The Pneumobil competition [17] is an excellent example of managing a product development project with heterogeneous teams in all aspects. Students give the core of the project, and they have to work with professors, technical staff, and corporate experts. Success can be measured in speed and distance, but there is more behind performance than technical excellence. A successful Pneumobil team is a model of a successful project team. Actually, a Pneumobil project is a great simulation medium for product development projects. Beyond the deliverables, monitoring the teamwork and project management practices offers lessons for both the companies and the university. Understanding the motivations, communication, and collaboration can lead to an improved project management method as well as it will highlight the lack of education.

Our research contributes to exploring the differences in the approach of product development experts and students. The scope of the pilot study is limited to the judgment of engineering and business management students on the success factors of product development projects.

2 Product Development Process Approaches

The managerial framework of product development projects is usually moved from the traditional waterfall approach to iterative or incremental ones [3] [10] [18]. These approaches allow the redesign of the following steps depending on the progress and new information available. Integrated product development models reflect the accelerating changes and follow the opportunities of new technologies.

Olsson [19] published a pioneer model of integrated product development that emphasized project orientation and the role of teamwork in product development. Development phases (investigation of need, product principle, product design, product preparation, and execution) are derived from the recognition of needs and defined by professions, including marketing, operations, or finance. Andreasen and Hein [20] focused on the parallel design of product versions that were available through computer-aided design. Time-saving was available in a stable market environment. Meerkamm [21] carried on the idea of integration that required flexible organization and a new approach to design, the design for the product life cycle. The goal is to create a product that meets the requirements for the first try and is finished on schedule according to the plans. The approach corresponds to a narrower interpretation of the project success described by the iron triangle. Among the characteristic models defined by the PMI [3], the iterative life cycle is its counterpart. Ottoson [22] emphasized flexibility through framework thinking rather than traditional processes. This approach allowed a more effective response to the environmental changes for the project team and the organization. Incremental characteristics [3] can be found in the model.

Human competencies have emerged in the model of Ehrlenspiel [23], who forced exploiting human resources and promoted motivation. The suggestions are in line with the even more popular lean approaches [24] [25]. Magdeburg model [26] emphasized human-centered thinking and enhanced communication again. The collaboration of human and organizational resources with technology and methodology is justified by these models. These models present the processes as dynamic networks.

3 Expert Opinions about Project Success Factors in Product Development Projects

Project success and failure have extensive industry-specific literature with some elements of common thinking [27]. The emphasis has gradually moved from the success of project management (the 1960s–1980s) to project and product success (the 1980s–2000s). Portfolio and program [28] success has been appreciated recently [5]. Emam and Koru [29] identified the top project cancellation reasons. Although the investigations were focused on IT projects, similar problems are mentioned in other industries:

- ☐ senior management is not sufficiently involved,
- ☐ too many requirements and scope changes,
- ☐ lack of necessary management skills,
- ☐ over budget,
- ☐ lack of necessary technical skills,
- ☐ no more need for the system to be developed,
- ☐ over schedule,
- ☐ technology is too new; it does not work as expected,
- ☐ insufficient staff,
- ☐ critical quality problems with software,
- ☐ end users are not sufficiently involved.

A former analysis of the authors focused on the product development experts' judgment on the success factors of project management [9] [30] [31]. These questions provided the initial for a student-level survey designed to compare the two groups' opinions. The survey included success factors related to the followings:

- ☐ regulation,
- ☐ information management,
- ☐ the collaboration of the project team,
- ☐ the focus of the project manager.

Figure 2 shows the order of the success factors based on the responses of 112 experts based on [9].

The research found a high-level of agreement on a clearly written set of project goals, as a specification or scope, is really the most important need for projects. Justifying the findings of Ehrlenspiel and Meerkamm [32], the need for cooperation within the project team was confirmed, as well as the active focus of the project manager on the project team, regulation, and collaboration topics. The supporting role of knowledge management and information technologies is confirmed by the results [31].

Figure 2
Importance of product development success factors by experts (five-point scale, the higher values show more important items)

4 Research Design

4.1 Research Goal

The research aims to explore the students' opinions about product development projects' success factors and compare the results to expert opinions for supporting further development of product development. Additionally, the results may contribute to building harmonious partnerships between corporations and higher education institutions. The results are expected to serve the improvement of project management education by highlighting the lack of coherence in the knowledge level. Since the research has an explorative characteristic and we consider it a pilot study, no statistical hypothesis is formulated. The research questions are as follows:

- ☐ What factors do students consider relevant to product development project success?
- ☐ What are the differences in the perception of success factors of the product development project between the present and future generations?

4.2 Survey Design

A voluntary online survey was designed to collect information about students' approaches and opinions about product development processes. Data processing was anonymous; the mean values and the distribution of the responses give the basis

of the conclusions. This study processes a highlighted question group about the success factors. The question is formulated as 'How critical do you consider the following factors for the success of product development projects?'. The respondents were asked to indicate if there are any shortcomings in the area that will impair the project's success. A five-point scale was used for the evaluation; higher values were asked to be marked if the respondent evaluated it harmful to the project if the item was corrupted or missing. The list of success factors investigated is based on the former survey among experts [9]:

☐ Clear, written project goal,

☐ Cooperation within the project team,

☐ Active focus on project team by the project manager,

☐ Project feedback meeting, collection of project lessons,

☐ Regular project meetings,

☐ Active focus on project deliverables by the project manager,

☐ Available written internal standards and regulation,

☐ Involvement of manufacturing experts into the product development project,

☐ Compliance with previously defined objectives and targets,

☐ Regular inspection of written standards.

4.3 Research Sample

The sample includes 155 responses from engineering and business students at the University of Miskolc in Hungary.

The data collection method was convenient, and the representativeness of the sample was not checked. Sample characteristics are summarized in Table 1.

Table 1
Students' sample composition

		Percent
Gender	Female	47.7%
	Male	52.3%
Study type	full-time	40.6%
	part-time	59.4
Study level	bachelor	51.6%
	master	48.4%

4.4 Analysis Methods

The survey was designed for statistical analysis considering the instructions of [33] and [34]. According to the research goal and purpose, the focus of the study is exploring the characteristic patterns among the respondents. Although the measurement level of most questions allows the analysis of the distribution [35], the students' responses are presented per question also by the mean values considering its limitations and bias [36]. The reason for using the mean values is a simplified presentation of the results; variance analysis uses nonparametric methods procedure [37] [38].

The study presents the differences in the evaluations by students and experts; the related rank order is based on the mean values. The paper includes a two-dimensional visualization of the results. The visualization logic importance-performance analysis [39] [40] is applied to explore the agreement level between students and experts.

Exploring the patterns of opinions was performed by cluster analysis [41]. The two-step method of IBM SPSS [37] was selected for the database.

The statistical analyses were performed at a 95% confidence level. Data analysis used Microsoft Excel and IBM SPSS.

5 Results and Discussion

5.1 Students' Assessment Results

Available standards and regulations are considered the most important success factor of a product development project in the sense that students feel the project implementation is unattainable in the absence of this. Moreover, the following items in the ranking also belong to regulatory issues (Figure 3). The project management activity and regular project meetings are at the end of the relative order.

Figure 3
Critical factors by students' evaluation (5-point scale, the higher values show more critical items)

Gender (female, male), study level (bachelor, master), and study type (full-time, part-time) were used as grouping factors of the analysis. Females' evaluation shows higher values than males, and part-time students' evaluation exceeds full-time ones for each item. In the mirror of efforts to increase the share of female engineers [42], a detailed analysis is worthwhile in this field. The evaluation by study level does not show a clear pattern, and the values are close to each other. However, the Kruskal-Wallis analysis of variance shows significant differences in limited cases:

- ☐ Active focus on project deliverables by the project manager (x_{male}=2.28, x_{female}=2.78, Kruskal-Wallis H=3.889, d_f =1, sig.=0.049), by gender
- ☐ Compliance with previously defined objectives and targets ($x_{full\text{-}time}$=2.32, $x_{part\text{-}time}$=2.88, Kruskal-Wallis H=8.953, d_f =1, sig.=0.003), by study type
- ☐ Active focus on project deliverables by the project manager ($x_{full\text{-}time}$=2.19, $x_{part\text{-}time}$=2.75, Kruskal-Wallis H=5.450, d_f=1, sig.=0.020), by study type

The values of the item-level evaluations are close to each other, suggesting a kind of uncertainty in students' opinions about the success of product development projects. As extensive experience in project management cannot be expected among the respondents, this is not reprehensible. Despite the limitations of the results, the relative order on the sample can describe the assumptions and imaginations that are valuable information for developing related curricula and preparing the companies for what they will face when hiring beginners. According to the survey experience, the goals, standards, and regulations are trusted.

5.2 Results of Cluster Analysis

The student's evaluations of the product development project success factors were used to establish clusters in order to explore patterns. A dimension reduction was necessary since the nonparametric correlation coefficient values between the question are high and significant in each case (Spearman's Rho values are between 0.407 and 0.909). Based on the four factors explored by principal component analysis with Varimax rotation, the two-Step Cluster analysis of the IBM SPSS software allowed three clusters with a fair explanatory power (average silhouette measure of cohesion and separation is 0.3). Figure 4 compares the mean values of the responses by the clusters.

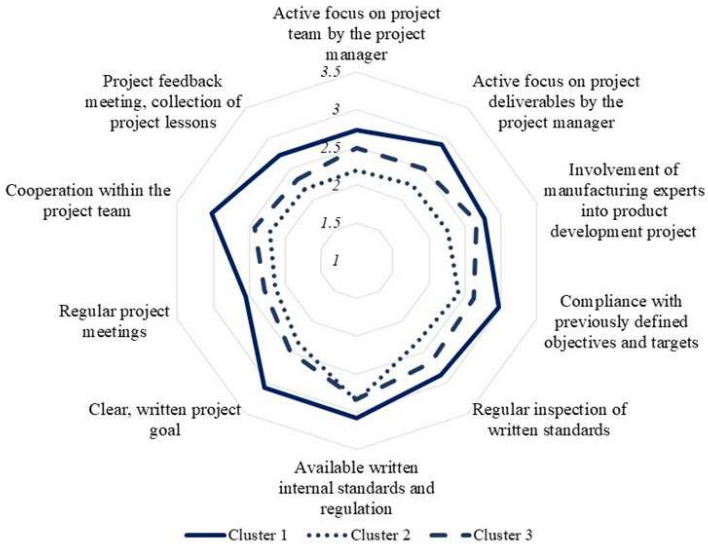


Figure 4

Cluster characteristics (5-point scale, mean values)

Cluster 1 (31.6% of the respondents) includes students with the highest ratings. They keep the factor more critical to project success than others. Outstanding values are found in the factors of project team cooperation, having a clear project goal, and the active focus of the project management on the deliverables. Cluster 2 (40.6% of the respondents) represents the middle way. In the case of Cluster 3 (27.7% of the respondents) show, the role of standards and regulation is dominant compared to other factors. They evaluated the failures in the listed success factors as less harmful than other clusters. The results confirm the general trust of students in the regulated procedures and the subordinate expectations about cooperation within the project team.

Cluster membership is checked for the grouping factors used in the survey by cross-tabulation, but no significant differences are found by gender, study level, or study type.

5.3 Comparing Student and Expert Assessment

The order of the mean values is shown in Table 2. The difference between expert and student evaluation is remarkable. The correlation coefficient between the rankings is at a moderate negative (ρ =-0.37) level, which suggests a lack of agreement. Figure 5 gives a visual representation and grouping of the results by the ranking orders.

Table 2
The ranking order of the evaluations

	Expert	Student
Clear, written project goal	10. the most important	8.
Cooperation within the project team	9.	5.
Active focus on project team by the project manager	8.	3.
Project feedback meeting, collection of project lessons	7.	2.
Regular project meetings	6.	1. the least harmful if failed
Active focus on project deliverables by the project manager	5.	4.
Available written internal standards and regulation	4.	10. the most harmful if failed
Involvement of manufacturing experts into product development project	3.	6.
Compliance with previously defined objectives and targets	2.	9.
Regular inspection of written standards	1. the least important	7.

There are only two items that are considered in the similar evaluation quarter by the experts and the students. The quarter in Figure 5 evaluated as important by the experts and less critical by the students include factors that cover the soft side of project management, such as the role of meetings or the personal focus of the project manager on the team as well as cooperation. Cooperation within the project team also belongs to this quarter. In addition, hard elements such as the role of written project goals, regulations, and inspection of the standards show just the opposite. These factors are considered more critical by the students than by the experts.

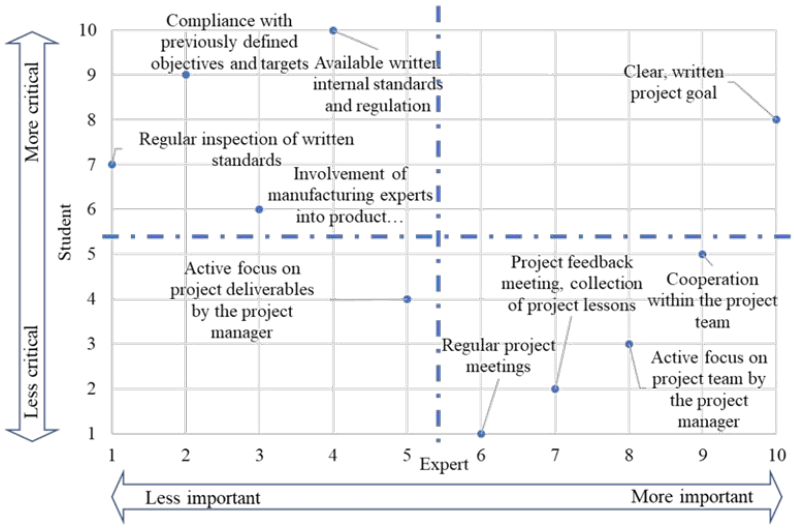


Figure 5
Visualization of joint results among expert and student

Conclusions

The unpredictable environment due to the accelerating technological and social changes force rethinking business processes. The pressure on product development can be tracked well in the improvement of the development models. The reduced market lead time of a new product requires additional resources from corporations that can be supported by involving higher education students. Sustaining team effectiveness needs understanding the similarities and differences in attitudes between the students and the corporate member.

According to the first research question about the relevant success factors considered by students, the belief in hard factors as standards as planning can be highlighted. It can be concluded that students imagine product development projects as manageable in a predictable way. The second research question aimed to compare the approach to project work between practicing professionals, and future project participants can provide profitable information. The survey results show remarkable differences between the opinions of the two groups. Students recognize hard factors of project success as more relevant, while experts value soft coordination tools as decisive. The experience of the study indicates the necessity for enhancing management education among engineering students (e.g., leadership, communication, problem-solving). The findings lead to a theoretical implication about the further directions of product development models. The option for collaboration is incorporated into recent methodologies, but it primarily aims for internal stakeholders, involving customers or procurement partners. Refinements should address students as future engineers and higher education institutions as knowledge sources.

Although the evaluation of the results is limited due to the sample composition, and the survey covers a pilot study of the problems, attention must be drawn to the improvement opportunities of engineering and management youth education. Product development projects involve a great variety of professional knowledge at a strenuous work pace. A common understanding of the values that drive project management is inevitable for assuring success at any interpretation level. Verifying the findings in other areas of specialization is necessary for generalizing the results.

Beyond improving the project management skills of students in general, the practical implication of the study is enhancing the cooperative programs between corporations and higher education institutions. Special projects, like Pneumobil projects, have a relevant student attendance and offer excellent opportunities for testing targeted project management training programs at low business risk.

References

[1] OECD/Eurostat. Oslo Manual 2018. Guidelines for collecting, reporting and using data on innovation, 4th Edition, The measurement of scientific, technological and innovation activities, Paris-Luxembur OECD Publishing-Eurostat, 2018, 254p.

[2] Deutsch, N., Mészáros, T., and Szabó, L. Modelling the development of strategic management. In Proceedings 31st European Conference on Modelling and Simulation: ECMS 2017. Edited by Zita Zoltay-Paprika, Péter Horák, Kata Váradi, Péter T., Zwierczyk, Ágnesd Vidovics-Dancs and Péter Rádics, Regensburg, European Council for Modelling and Simulation, 2017, pp. 67-73

[3] PMI. Agile practice guide, Newtown Square, Project Management Institute, 2017, 210p.

[4] Baccarini, D. The logical framework method for defining project success, Project Management Journal, Vol. 30, No. 4, 1999, pp. 25-32

[5] Ika, L. A. Project success as a topic in project management journals, Project Management Journal, Vol. 40, No. 4, 2009, p. 619

[6] Pollack, J., Helm, J., Adler, D. What is the iron triangle, and how has it changed? International Journal of Managing Projects in Business, Vol. 11, No. 4, 2018, pp. 527-547

[7] Görög, M. Project management in organizations, Budapest, Panem, 2019, 601p.

[8] Verzuh, E. The fast forward MBA in project management, 5th ed., Hoboken: John Wiley & Sons, 2015, 496p.

[9] Soltész, L., Berényi, L. Success factors in product development projects: expert opinion, Journal of Physics: Conference Series, Vol. 1935, 2021, paper 012001

[10] Vajna, S. Integrated design engineering, Berlin-Heidelberg, Springer Vieweg, 2014, 613p.

[11] Pahl, G., Beitz, W., Feldhusen, J., Grote, K. H. Engineering Design. A systematic approach, London, Springer, 2007, 617p.

[12] Bai, H., Arcak, M., Wen, J. Cooperative control design. A systematic, passivity-based approach, New York, Springer, 2011, 209p.

[13] Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., Leifer, L. J. Engineering design thinking, teaching, and learning, Journal of Engineering Education, Vol. 94, No. 1, 2005, pp. 103-120

[14] Ciric, D., Lalic, B., Gracanin, D., Palcic, I., Zivlak, N. Agile project Management in New Product Development and Innovation Processes: Challenges and Benefits Beyond Software Domain, IEEE International Symposium on Innovation and Entrepreneurship (TEMS-ISIE), Beijing, China, 2018, 9p.

[15] Lindlöf, L., Furuhjelm, J. Agile beyond software - A study of a large scale agile initiative, Proceedings of the DESIGN 2018 15th International Design Conference, Dubrovnik, Croatia, 2018, pp. 2055-2062

[16] Varl, M., Duhovnik, J., Tavčar, J. Agile product development process transformation to support advanced one-of-a-kind manufacturing, International Journal of Computer Integrated Manufacturing, Vol. 33, No. 6, 2020, pp. 590-608

[17] Kelemen, L. 10 years of Pneumobil competition at the University of Miskolc, Gép Journal, Vol. 68, No. 4, 2017, pp. 21-24

[18] Shrivastava, A., Jaggi, I., Katoch, N., Gupta, D., Gupta, S. A systematic review on extreme programming, Journal of Physics: Conference Series, Vol. 1969, 2021, paper 012046

[19] Olsson, F. Integrated Product Development – Working Model. Series 21 Product Development. Stockholm, Sveriges Mekanförbundet, 1985

[20] Andreasen, M. M., Hein. L. Integrated product development. Berlin-Heidelberg, Springer, 1987

[21] Meerkamm, Harald. 1995. Integrated product development in the area of conflict between cost, time and quality management. VDI-Jahrbuch 95, Düsseldorf, VDI-Verlag

[22] Ottosson, S. Dynamic product development: Findings from participating action research in a fast new product development process, Journal of Engineering Design, Vol. 7, No. 2, 1996, pp. 151-169

[23] Ehrlenspiel, K. Integrated product creation organization - Methods - Tools. In Milberg, J., Competitive factor time in production companies, Berlin-Heidelberg, Springer, 1991, pp. 113-131

[24] Likert, J. K. The Toyota Way: 14 management principles from the World's greatest manufacturer, New York, McGraw Hill, 2004, 330p.

[25] Womack, J. P., Jones, D. T. Lean thinking: Banish waste and create wealth in your corporation. Revised and updated, New York, Free Press, 2003, 396p.

[26] Burchardt, C. An extended concept for integrated product development. In Vajna, S., Integrated Product Development. Vol. 3, Magdeburg, Lehrstuhl für Maschinenbauinformatik, 2001, 208p.

[27] Müller, R., Judgev, K., Critical success factors in projects: Pinto, Slevin, and Prescott – the elucidation of project success, International Journal of Managing Projects in Business, Vol. 5, No, 4, 2012, pp. 757-775

[28] PMI. A Guide to the Project Management Body of Knowledge (PMBOK Guide), 6th ed., Newtown Square, Project Management Institute, 2017, 756p.

[29] Emam, K. E., Koru, G. A. A Replicated survey of IT software project failures. IEEE Software, Vol. 25, No. 5, 2008, pp. 84-90

[30] Soltész, L., Berényi, L., Kamondi, L. Analysis and assessment of the product development process, Gép Journal, Vol. 71, No. 4, 2020, pp. 67-71

[31] Venczel, T. B., Berényi, L., Hriczó, K. Project management success factors, Journal of Physics: Conference Series, Vol. 1935, 2021, paper 012005

[32] Ehrlenspiel, K., Meerkamm, H. Integrated product development: thought processes, use of methods, collaboration, München, Carl Hanser Verlag GmbH & Co., 2009, 826p.

[33] Babbie, E. The Practice of Social Research, 15th ed., Boston: Cengage, 2020, 592p.

[34] Creswell, J. W., Creswell, J. D. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 5th ed., Thousand Oaks, SAGE, 2017, 304p.

[35] Kindler, J., Papp. O. Complex systems analysis: comparison methods, Budapest, Műszaki Könyvkiadó, 1977, 262p.

[36] Lalla, M. Fundamental characteristics and statistical analysis of ordinal variables: A review, Vol. 51, 2017, pp. 435-458

[37] Pallant, J. SPSS Survival Manual: A Step by step guide to data analysis using IBM SPSS, 7th ed., London, Open University Press, 2020, 378p.

[38] Higgins, J. J. An introduction to modern nonparametric statistics, London, Duxbury Press, 2003, 384p.

[39] Martilla, J. A., James, J. C. Importance-Performance Analysis, Journal of Marketing, Vol. 41, No. 4, 1977, pp. 77-79

[40] Lin, S. P., Chan, Y. H., Tsai, M. C. A transformation function corresponding to IPA and gap analysis, Total Quality Management & Business Excellence, Vol. 20, No. 8, 2009, pp. 829-846

[41] Verma, J. P. Data Analysis in Management with SPSS Software, New Delphi, Springer, 504p.

[42] Zsótér, B., Gál, J. Increasing the number of female students in the Faculty of Engineering, Jelenkori társadalmi és gazdasági folyamatok, Vol. 14, No. 1, 2019, pp. 199-203

Design, Production and Analysis of Cylindrical, Coaxial and Alternating Current Shunts, using ANSYS

Igor Štambuk*, Jure Konjevod, Roman Malarić****

*Centre for Defense and Strategic Studies ¨Janko Bobetko¨, Croatian Defense Academy "Dr. Franjo Tuđman", Ministry of Defense of the Republic of Croatia, Ilica 256 b, 10000 Zagreb, Croatia, istambuk@morh.hr

**Department of Electrical Engineering Fundamentals and Measurements, Faculty of electrical engineering and computing, University of Zagreb, 10000 Zagreb, Croatia, jure.konjevod@fer.hr, roman.malaric@fer.hr

Abstract: This paper describes a new approach for designing with the ANSYS program, including cylindrical, coaxial and alternating current shunts, designed for currents from 10 A to 100 A, which are used as new work standards of the National Research Council (NRC), in the frequency range up to 100 kHz. The heating factor, is one of the more important issues, when determining the dimensions, i.e., the maximum current for which the shunt is intended. Using the ANSYS software package, a model for the current shunt is designed and the distribution of the temperature flux, of the coaxial shunt is analyzed. Further, the main stages of the 3D modeling of the temperature field are considered. The theoretical temperatures of the cylindrical shunt obtained by simulation, are compared with the experimentally determined temperature values, of the resistive shunt, of the same dimensions and characteristics, as carried-out in the laboratory, using the same currents.

Keywords: resistive coaxial shunt; ANSYS; 3Dmodeling; electrical power; metrology; resistive current transducer

1 Introduction

Today, due to the global energy crisis, power measurement is gaining more and more importance. The measurement of voltage and current magnitudes, provide a direct way of controlling the electric voltage and current in a circuit and the measurement of electric current can be achieved directly, by observing the effects of the current itself, or through the voltage drop across a well-defined impedance, which is considered a more stable and convenient method.

Current shunts, as one of the most common devices for measuring current, work as a passive device. Due to their excellent performance from DC level and up to high frequencies, even to 1 MHz, they are usually designed for the needs of AC electrical circuits [1] [2]. The main characteristics of current shunts are long-term stability, temperature and power coefficients (TCR and PCR), AC-DC difference and phase angle [3]. To create a functional high-current and high-frequency shunt, it is important to minimize phase error. The phase angle measurement error of current shunts is elaborated in [4]. The construction of the shunt can be performed using different techniques and methods, for example in the form of a foil and/or coaxial shunt, and the most common is the cage type. Such cage coaxial shunts are usually thoroughly characterized as presented in [5] and [6] respectively. Measurement setup and measurement method for AC-DC difference determination of the cage current shunts is presented in [7] and [8] respectively. However, in recent years the need for foil shunts has reemerged due to their exceptional conductivity at high frequencies. Design and fabrication of foil shunts in BEV, Austria with preliminary experimental results is presented in [9] while AC-DC difference of BEV foil shunts is presented in [10].

In the previous work, the emphasis was on measurement precision [11], designing a system for precise measurement of resistance etalons [11], and making elements that can supply the measuring system with a sufficient amount of stable current [12]. Significant improvements in NRC AC-DC differential current transfer measurement capability [2], intended mainly for applications in the frequency range from 10 Hz to 100 kHz, have been presented. The first step in the construction of the new foil shunt 100 A, 100 kHz [13] was also presented, and the parameters of the foil construction were defined, and based on the analysis and the obtained results, a method of determining the dimensions was developed. Thus, a newly designed shunt is proposed.

Different techniques and models are used in today's endeavors to have as accurate predictions as possible such as described in [14-17]. The ANSYS simulation system was also used in many tests [18-21] of structural durability, electrical and thermal properties, but with the development of application software, the ANSYS system offers us some new possibilities that are described in this paper. First of all, these are the possibilities of testing in the simulation before the shunt is produced and determining the behavior of the shunt when high alternating currents flow through it.

This paper presents the modeling of the shunt used in measuring systems with alternating current by the simulation system ANSYS [22]. Algorithms are explained gradually in chapter number two, through the presentation of individual stages of model development, from modeling in the SpaceClaim application (which is an integral part of the ANSYS program), through the process of defining the resolution and details of the analysis itself (Accurate Meshing), to building the model. After entering the Thermal electric model, and defining Steady-State Electric Conductions, where the voltage, current or heat that we supply to the model is

defined, along with the temperature of the environment in which the analysis will take place. In the third chapter, the results of analyzes that can be performed with the existing designed model are presented, so the electric field on the input side of the shunt and Total heat flow on the surface of the model are presented.

In order to analyze the results obtained with the simulation system ANSYS, [21] an AC shunt was created. The shunt is gradually heated over a period of ten minutes, and the temperature on the shunt surface is compared with the experimentally obtained data from ANSYS, in which the same heating conditions are simulated. The obtained data are connected, analyzed and given in the form of a conclusion.

2 Current Coaxial Shunt Design

In order to choose the appropriate simulation program, we tested several versions of the ANSYS Simulation & Design applications. We are looking for a model that in a relatively short time can show those parameters in the process of testing the characteristics of current shunts that would take a lot of time, i.e., it would be necessary to create a physical model of the current shunt and feed it with alternating current from a generator, and measure the processes mostly with expensive equipment in the laboratory. The capabilities of the tested simulation systems vary both in the area of design and in the area of calculation of the required display sizes and resolutions. One such designed shunt model developed in ANSYS version 17.0 is shown in Figure 1.

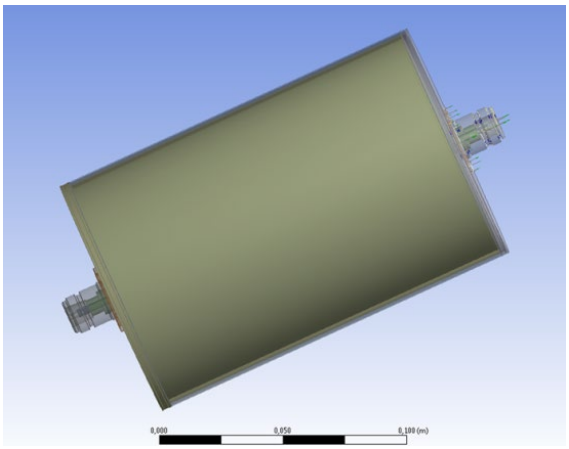


Figure 1
Shunt model created in the ANSYS 17.0 program

2.1. Modeling in the SpaceClaim Application

The final program selection was ANSYS version 2019. After defining the dimensions of the shunt, we started creating the model. Modeling is done in ANSYS SpaceClaim, which contains tools to speed up geometry preparation for modeling with 3D responses but is faster than complex traditional CAD systems. Geometry preparation is quite fast, as is shunt parameter entry, thanks to development tools to accelerate geometry preparation, which achieves a faster transition to simulation while eliminating delays between design teams.

Compared to modeling in ANSYS version 17.0, where individual parts of the assembly are added separately, SpaceClaim offers easy 3D modeling and provides tools for speeding up geometry preparation and creating much more complex 3D shapes in one part, as well as quick repairs of the model itself. Models created in the SpaceClaim application are compatible and customized with ANSYS version 19.0.

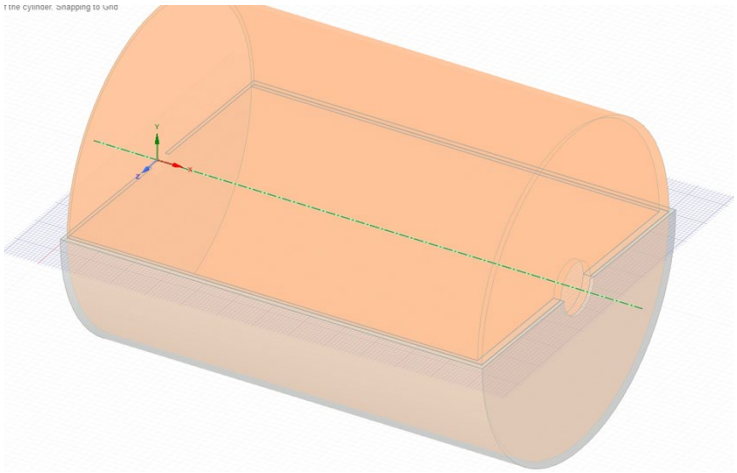


Figure 2
Designing a shunt in the ANSYS SpaceClaim program

Figure 2 shows the process of designing a shunt in such a way that layers of copper and manganin are designed around the PVC core, first those in the current part of the circuit. Between them, a layer is designed that contains properties of the Kaplan tape that was used as insulation between the conductive layers of the shunt.

Thermophysical properties of the manganin material of interest are given in Table 1. [23] [24] Manganin is a trademarked name for an alloy of typically 84.2% copper, 12.1% manganese, and 3.7% nickel.

All assembly segments are assigned material and connection properties. The following materials were used in this work: PVC, copper, manganin and Kaplan tape. In the process of creating the model, we also used other simulation programs

and compared the design results. In some applications, we have noticed defects related to performances with tape thickness in the process of connecting individual parts of the module, i.e., the inability to follow the line (broken line) of the model.

Table 1
Manganin thermophysical properties

Electrical Resistivity at 20 °C	0.43 μΩ/m
Thermal Conductivity at 20 °C	22 W/mK
Specific Heat at 20 °C	0.41 J/gK
Density at 20 °C	8.4 g/cm^3
Tensile Strength at 20 °C	390 MPa
Melting Point	960 °C

2.2. Efficient Solutions from Accurate Meshing

One of the essential segments of shunt design is the selection of resolution for Accurate Meshing. ANSYS provides automated software that produces a suitable mesh for accurate, efficient multi-physics solutions. In this case, due to extremely thin layers, systematic automatic meshing of a highly constructed mesh with the necessary resolution to correctly capture solution gradients is required for reliable results. However, care should be taken that too high a resolution, increasing the amount of data, does not lead to system congestion and prevents the simulation itself from working properly.

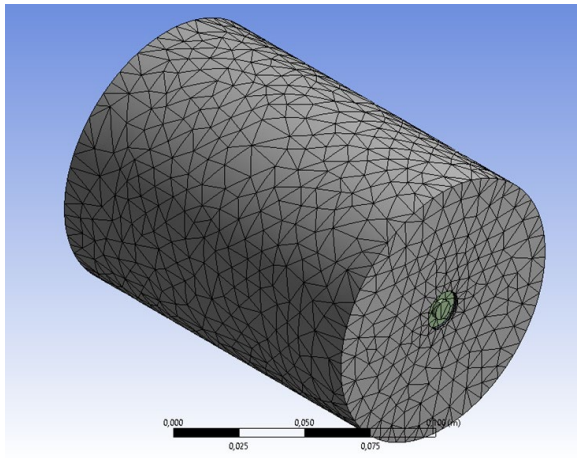


Figure 3

Choosing a resolution for Accurate Meshing

3 Measurement Results Analysis

In the process of building the model, after entering the Thermal electric model, we define Steady-State Electric Conductions, where we define the voltage, current or heat that we supply to the model, as well as the temperature of the environment in which the analysis will take place. Next, we define the information about the ¨Solution¨ that we want to get from the performed analysis, for example it is ¨Total electric field intensity, Total current density, Total heat flows, etc.). In the model, one can experiment with several input parameters, based on which is selected and which type of analysis is run for certain settings. As an example, analysis of the total electric field distribution and Total heat Flux are shown below.

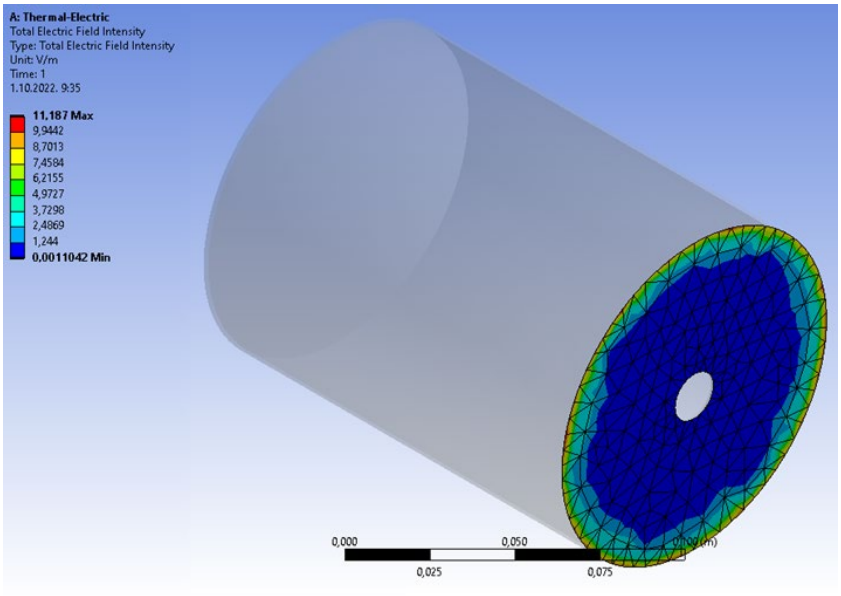


Figure 4
Distribution of the electric field on the input side of the shunt

Figure 4 shows the distribution of the electric field on the input side of the shunt, in the simulation program ANSYS, after applying a voltage of 10 V to the input of the shunt. It can be seen that the distribution of the electric field is strongest at the edges of the wall itself, which is expected due to the layer of manganin wrapped around the PVC core, which causes a stronger field with its resistance.

If we make an analysis in the ¨Thermal-Electric¨ area, Total heat Flux (W/m^2), it can be seen that the greatest heating is inside the shunt and in the part of the manganin, which due to its higher resistance heats up faster than other elements of the circuit.

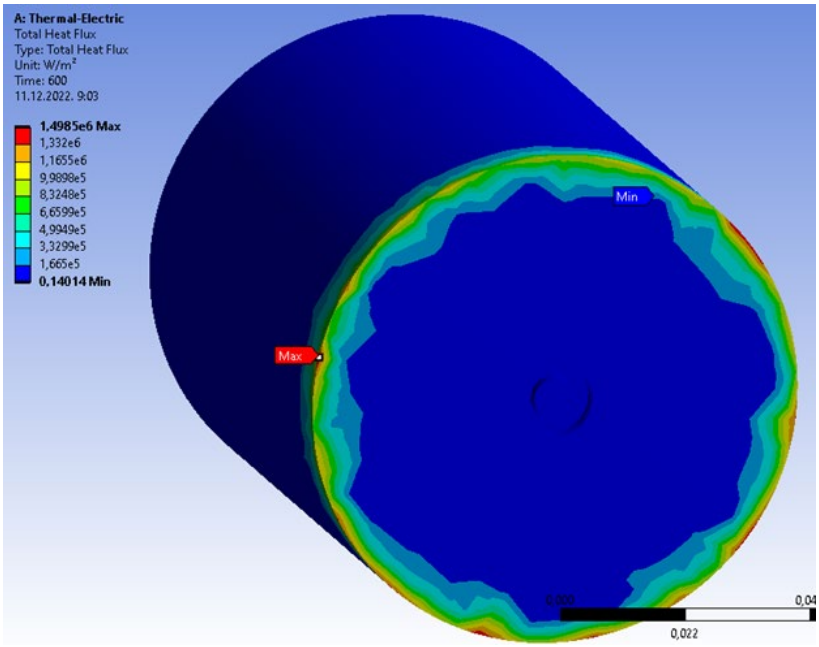


Figure 5

Total heat flow (W/m^2) on the surface of the model

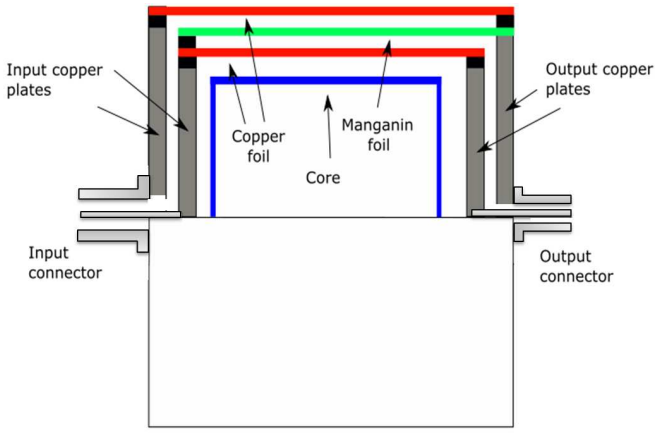


Figure 6

Simplified schematic of a tested cylindrical coaxial shunt

In order to compare the obtained results, we used a thermal image of a previously made and tested shunt. The tested resistive element is a cylinder made of thin manganin film with a thickness of 0.022 mm. Four circular copper plates, made of copper with a thickness of 5 mm, form parts of the input current and output voltage lines. The cylinder of the return current and the cylinder of the high-voltage line are

made of a thin copper foil with a thickness of 0.1 mm. Copper cylinders are located outside and inside the manganin cylinder. These three coaxial cylinders, molded on a glass fiber epoxy core, are mutually insulated and, if necessary, connected only at the inlet and outlet plates. The basic diagram of the shunt is shown in Figure 6.

3.1. Measurement of Resistance Shunt Heating in Laboratory Conditions

The measurement setup for resistance shunt heating in laboratory conditions determination is consisted of the function generator RIGOL DG4062, power amplifier TOELLNER TOE 7621 and tested current shunt. The function generator provides an input signal for the amplifier. Output of the applied amplifier is then connected to the input connector of the current shunt. The measurement scheme and the appearance of the tested shunt are shown in Figure 7.

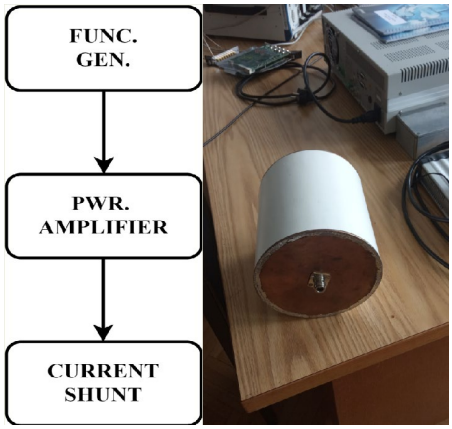


Figure 7
Measurement scheme and tested current shunt

Coaxial shunts were designed in such a way that each Vishay resistor is not loaded more than 30% of their power ratings. In such a way shunt will not be powered at maximum power ratings and will not overheat. Thus, for most practical applications, correction of measurement results due to self-heating, will not have to be performed, since the Vishay resistors have a low temperature coefficient. In the same way foil shunts will be operated at lower currents than possible even though Manganese foil has temperature coefficients in the range of several ppm/°C to avoid performing temperature coefficient corrections.

The main equation which is widely used for current measured with the current shunt is [25]:

$$I = \frac{U \cdot (1+ \delta_A + \delta_{sv} + \delta_{res} + \delta_{le})}{R_s \cdot (1+ \delta_{Rs} + \delta_{drift} + \delta_{ac-dc} + \delta_{tcr} + \delta_{pwc})} \quad (1)$$

, where U is measured ac voltage across the shunt, R_s is the DC resistance of the current shunt. Coefficients δ with related indexes in numerator represents corrections due to the repeatability of the ac voltage measurement, correction due to ac voltmeter deviation and due to the resolution of the AC voltmeter and correction due to the load effect on the reference AC voltmeter. Further corrections δ in denominator are due to the DC resistance measurement deviation, due to the drift, AC–DC difference, temperature and power coefficient of the shunt [25].

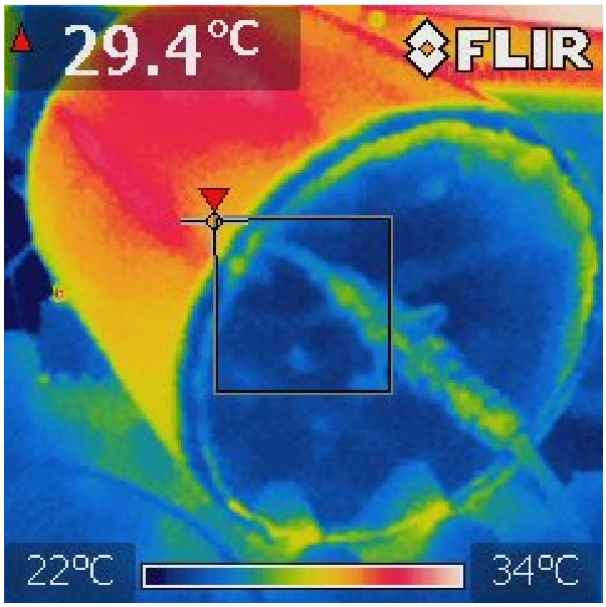


Figure 8
Image of shunt heating taken with FLIR thermal camera

An alternating current of 10A was connected to the shunt, and heating was monitored with a thermal FLIR camera. Fig 8 was created after 10 minutes of heating, and it clearly shows the degree of heating of individual parts of the shunt. The maximum temperature measured was on the formwork, that is, the outer layer of the resistive shunt, and it was 29.4 °C. The image shows a weaker heating of the side copper plates, due to their thickness, and a greater heating in the area of thin layers of conductive materials, more in the case of magnesium and less in the case of copper.

The disadvantage of this method of measuring with a thermal camera is the possibility of measuring temperature heating by layers separately, that is, from the image we only see the surface temperature, which in this case was caused by the heating of the layer below.

3.2. Simulation in the ANSYS Application and Analysis of the Results

In the simulation, we set identical parameters, the input alternating current of 10 A, and the time in which the results will be monitored of 10 min. The heating process itself can be viewed in a short video that the application offers. For the current, it is necessary to set it to be alternating with a frequency of 50 Hz. The adjustment is done in the ¨mechanical¨ settings, and on the magnitudes, you should click on the arrow to the right and select function.

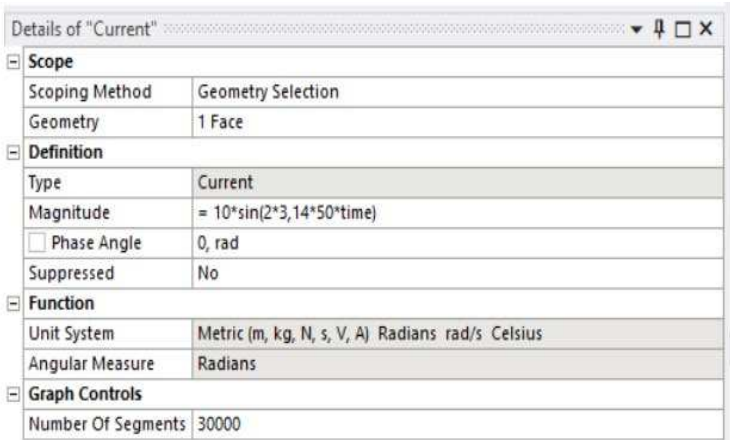


Figure 9

Settings for alternating current in the ANSYS application

If we compare the results, it is evident that the greatest heating is in the part of the manganin thin film, which has the greatest resistance and defines the total resistance of the shunt. It has been also seen that the mentioned heating penetrates through the surface of the shunt in the form of additional heating of the thin copper film located on the surface. The temperature that develops on the surface of the shunt in the case of measurement with a thermal camera is 29-31 °C, while in the case of simulation it is also 29.5-30.5 °C. The maximum temperature of the shunt was measured with a thermal camera at 34 °C, while in the simulation the maximum measured temperature was 32.5 °C. The maximum temperature develops on the inner layer of Manganin. Since the front entrance walls of the shunt are made of relatively thick copper plates, there was no significant increase in temperature during the observed period. The process of heating the shunt due to the flow of current can also be seen in video form in the ANSYS simulation.

It is also noticeable that the measurement with a thermal camera more vividly evokes thermal breakthroughs, in certain segments and creates a more realistic visual image of the processes taking place, while the simulation in the ANSYS application provides more detailed information for all of the processes and values achieved in the procedure.

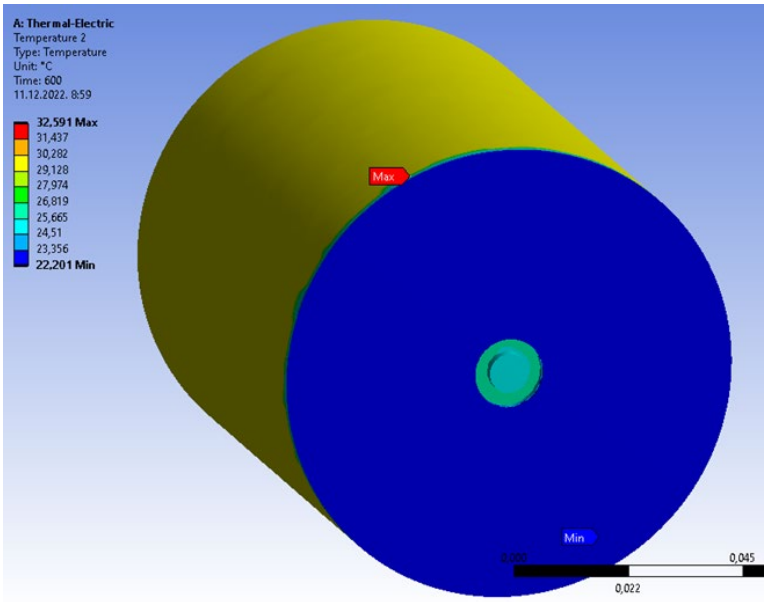


Figure 10

Temperature on the surface of the shunt after heating generated using the ANSYS application

Conclusions

Coaxial shunts are widely used, as part of electrical power control devices and power system testing equipment and for the precise measurements of alternating and direct currents, in wide dynamic and frequency ranges. This paper describes the design of the shunt, using a simulation program, as well as the results of the experimental performance and assessment of the influence of characteristics and dimensions, on the heating process, using the ANSYS software system. The analysis showed that the modeling of the thermal field and the calculated values of the heating temperature, of the shunt prototype, correspond, to a certain extent, to the values measured by the thermal camera. This research will make it possible to determine critical operating modes and identify the highest possible impulse current that will not lead to the destruction of the shunt.

References

[1] J. Webster: Current shunts in Wiley Encyclopedia of Electrical and Electronics Engineering, John Wiley and Sons, Inc, 1999

[2] C. Maxwell: A Treatise on Electricity and Magnetism, 3rd ed., Vol. 2, Oxford: Clarendon, 1892, pp. 68-73

[3] J. Konjevod, R. Malarić, M. Dadić, P. Mostarac and H. Hegeduš: Measurement of DC properties and relative humidity (RH) dependence of wideband AC current shunts, Elsevier Measurement 131 (2019) 1-6, 2019

[4] I. Budovsky: Measurement of phase angle errors of precision current shunts in the frequency range from 40 Hz to 200 kHz, IEEE Trans. Instr. Meas. 56 (2007) pp. 284-288

[5] G. Rietveld, J. H. N. van der Beek, E. Houtzager, DC characterization of AC current shunts for wideband power applications, IEEE Trans. Instr. Meas. 60 (2011) pp. 2191-2194

[6] B. Voljč, M. Lindič, B. Pinter, M. Kokalj, Z. Svetik, and R. Lapuh: Evaluation of a 100 a current shunt for the direct measurement of ac current, IEEE Transactions on Instrumentation and Measurement, Vol. 62, No. 6, 2013, pp. 1675-1680

[7] V. Novakova Zachovalova: AC-DC current transfer difference in cmi, Conference on Precision Electromagnetic Measurements (CPEM) 2008, pp. 362-363

[8] J. Konjevod, M. Dadić, R. Malarić, and I. Kunšt: Ac-dc transfer difference measurement of ac shunts, Journal of Energy, Vol. 67, No. 1, 2018

[9] P. S. Filipskil, M. Boeckerl, M. Garcocz: 20-A to 100-A AC-DC Coaxial Current Shunts for 100 kHz Frequency Range, Instrumentation and Measurement Technology Conference - IMTC 2007 Warsaw, Poland, May 1-3, 2007

[10] M. Garcocz, G. Heine Scheibenreiter, W. Waldmann: Progress on BEV foil shunts for AC-DC current transfer, Precision Electromagnetic Measurements (CPEM 2014), Rio de Janeiro, Brazil, 2014, pp. 272-273

[11] I. Štambuk; R. Malarić: Measurement system for precise comparison of low ohmic resistance standards, Measurement, 66 (2013) pp. 161-167, doi:10.1016/j.measurement.2015.02.011

[12] I. Štambuk; R. Malarić: Development of 9 A current source for precise resistance measurement method, Review of scientific instruments, 83 (2012), 106105; 106105-1 doi:10.1063/1.4759022

[13] P. Mostarac, T. Hrkač, R. Malarić, I. Kunšt: Development and frequency analysis of AC foil resistive current transducer – preliminary results, 3rd International Colloquium on Intelligent Grid Metrology (SMAGRIMET) online event, October 20-23, 2020

[14] N. Ngoc Son, L. The Vinh: Parameter estimation of photovoltaic model, using balancing composite motion optimization, Acta Polytechnica Hungarica, Vol. 19, No. 11, pp. 27-46

[15] E. Hedrea, R. Precup, R. Roman, E. M. Petriu: Tensor product-based model transformation approach to tower crane systems modeling, Asian Journal of Control, Vol. 23, No. 3, 2021, pp. 1313-1323

[16] S. M. Abramov, S. Travin, G. Duca, R. Precup: New opportunities model for monitoring, analyzing and forecasting the official statistics on coronavirus disease pandemic (Rom. J. Inf. Sci. Technol. 26(1), 2023, pp. 49-64

[17] I. Borlea, R. Precup, A. Borlea: Improvement of K-means cluster quality by post processing resulted clusters (Proc. Comput. Sci. 199, 2022, pp. 63-70

[18] E. V. Bedareva, P. S. Marinushkin, and L. I. Khudonogova: Calculation of coaxial shunts superheat temperature, MTT'2014, 2014

[19] M. El Khaldi, J. Bonis, A. Camara, L. Garolfi, A. Gonnin: Electromagnetic thermal, and structural analisys of Athomx RF gun using Ansys¨, Proceedings of IPAC2016, Busan, Korea, 2016

[20] M. Venczel, M. Steidl, A. Veress: Design Modifications and Thermal Analysis of Visco-Dampers for Extending Silicone Oil Durability, Acta Polytechnica Hungarica Vol. 18, No. 8, 2021. pp. 27-46

[21] H. Yıldız, E. Uzal, H. Çalık: An Analytical Solution of a Multi-Winding Coil Problem with a Magnetic Core in Spherical Coordinates, Acta Polytechnica Hungarica Vol. 18, No. 10, 2021, pp. 87-112

[22] Ansys, ANSYS, Inc, https://www.ansys.com, 2023

[23] Isabellenhuette: Manganin® manufacturer's online datasheet, https://www.spacematdb.com/spacemat/manudatasheets/MANGANIN-ISABELLENHUETTE-R.pdf. Accessed 28 Dec, 2022

[24] A. Schmon, K. Aziz, M. Luckabauer, G. Pottlacher: Thermophysical Properties of Manganin (Cu86Mn12Ni2) in the Solid and Liquid State, International Journal of Thermophysics, Volume 36, 2015, pp. 1618-1626

[25] B. Voljc, M. Lindic and R. Lapuh: Direct Measurement of AC Current by Measuring the Voltage Drop on the Coaxial Current Shunt, in IEEE Transactions on Instrumentation and Measurement, Vol. 58, No. 4, 2009, pp. 863-867, doi: 10.1109/TIM.2008.2007074

Different Approach of the Digital Transformation at SME

Ádám Béla Horváth

Óbuda University, Doctoral School for Safety and Security Sciences, Bécsi út 96/b, 1034 Budapest, Hungary; horvath.adam@kgk.uni-obuda.hu

Abstract: The spread of Industry 4.0 technological solutions is fundamentally transforming the value creation processes of for-profit organisations in technology-intensive and non technology-intensive industries alike. An analysis of the transition (digital transformation) from Industry 3.0 to Industry 4.0 requires understanding the various factors that come into play in the establishment of the ICT infrastructure and the inter-relatedness of the various infrastructural elements of Industry 4.0 and its previous generations. Since in Hungary, digital transformation processes were previously only assessed with qualitative means, a questionnaire-based survey was conducted among Hungarian profit-oriented for-profit organisations in 2018 with the involvement of 498 respondents. The aim of the research presented here is to show that the ICT infrastructure in the life of a business organisation is not simply the result of an internal decision, but is also influenced by external factors. Furthermore, the article shows that, on the one hand, a well-functioning ICT infrastructure is perceived by top management as an operational, tactical and strategic benefits that can motivate innovation, and on the other hand, it provides the technological background for innovation. A research model has been set up and validated by descriptive and inferential statistical procedures to process the data from the questionnaire responses. The result is a model that explains under which boundary conditions a supportive role of ICT infrastructure can be expected to stimulate further innovation in business organisations.

Keywords: digital transformation; enterprise architecture; Industry 4.0; Internet of Things; small- and medium enterprises; supply chain

1 Introduction

The sudden, recent – in the past decade – shift and development of business-oriented information and communications technology (ICT) solutions have fundamentally changed the value creation activities of for-profit organisations. These technological advancements (including, but not limited to: Internet of Things, Artificial Intelligence, smart robots, augmented reality, Big Data, Blockchain, etc.) are collectively referred to as Industry 4.0 technologies [1].

The extent and speed of this transformation is so great that the for-profit organisations are facing a technological paradigm shift [2, 3].

Although the impact of the regular use of Industry 4.0 technologies is noticeable on the for-profit organisations, industrial and national economy levels [4], the present research shall focus on the level of enterprises: the impact of Industry 4.0 on for-profit organisations – based on Porter's Value Chain [5] – has an effect both primary and secondary activities. Industry 4.0 solutions support business competitiveness in several ways: some of them are quantitative, while others are qualitative. The qualitative dimension can be understood as, on the one hand, the replacement of human labor reduces wage costs and, on the other hand, a lower error rate reduces material costs of production. The quantitative dimension means that value-creating processes can react more quickly to changes in the market and customer needs, thus achieving higher customer satisfaction [6]. A noticeable similar development can be seen also in the field of support processes. Through the example of controlling activities, the cited article [6] demonstrates that as a support activity, the analysis of a larger quantity and broader body of historical and current data allow for a more accurate and better data quality. The accessibility of the advanced analysis is more suitable for supporting the management in achieving the long-term, strategic goals [7].

On the sectoral level, it is worth making a distinction between two phenomena: firstly, the production of numerous – previously non-technology-intensive – industries can be revolutionized by Industry 4.0. The best example of this is agriculture [8, 9] – which is faced with significant challenges all across the world due to global warming – or logistics, where the challenge is precisely to reduce carbon dioxide emissions, one of the causes of global warming, through increased efficiency. Apart from revolutionizing production processes, it can also have significant impact on previously non-technology-intensive sectors since Industry 4.0 solutions enable previously unfeasible production innovations as seen in the development of smart garments in the different industries [8, 10]. By analogy, even in the case of industries that were previously considered technology-intensive, there is a noticeable tendency of Industry 4.0 solutions realizing previously “unimaginable” innovations [2, 8].

The spread of Industry 4.0 represents a significant advancement within organizations as well as in terms of international relations as it functions as a catalyst through the vertical and horizontal supply chain integration of various for-profit organisations, standardizing data flow and integrating various production systems across organizations [11]. This is of great significance in terms of industrial relations, for example, in the case of the agricultural and food industry, where the same product is taken through the supply chains of companies belonging to various industries (thus realizing the domino-effect) and the product intended for end use must comply with strict legal requirements [12]. At the same time, however, although some Hungarian SMEs are not familiar with the term Industry 4.0, they still use some of its elements [13] and the majority of SMEs

recognise the advantages of digitalization in terms of financial performance, operation and strategic performance. The use of Industry 4.0 elements amongst Hungarian SMEs were in more studies examined, showed that small and mid-size enterprises mostly employ Industry 4.0 elements in the fields of customer relations, management and administration, as well as logistics [14, 15].

As pointed out above, the definition of Industry 4.0 includes a number of recently developed technological innovations. Although some publications treat these technologies as equals [3], the IoT-technology must be considered as the most fundamental Industry 4.0 technology. This is simply because IoT-systems enable data collection through their sensors, which means this is the point where the real, physical world is converted into data, which is then forwarded to another device or server through a data transmission system and then executed in a partly or fully automated manner through a decision-making process based on Big Data-analysis and artificial intelligence. Regardless of whether the sensors provide structured (e.g. production data) or non-structured data [6, 8, 11], the cited publication shows that Industry 4.0 is inevitable based on the continuous, high-accuracy provision of data.

The research presented in this publication is motivated by the need to build a model based on data from Hungary that analyses what can motivate the notoriously – and as we have seen in the demographic characterisation – capital-poor Hungarian SMEs to invest in their ICT infrastructure and become open to innovation in Industry 4.0 solutions. In the second chapter the literature background of the research is presented. In the third chapter the questionnaire survey will be described on which the research presented here is based. In the fourth chapter, the theoretical model of the research and the related questions of the questionnaire will be presented, as well as the statistical analysis of the responses to the questionnaires. In the last chapter before the final one, the results of the statistical analysis are evaluated and finally conclusions are drawn.

2 Literature Review

Measuring the impact of Industry 4.0 technological solutions integrated into the value creation processes of companies on profit-oriented organizations is hindered by the fact that one must take stock of a number of – vastly different – technologies when striving for an exhaustive overview [16]. When examining the relevant qualitative studies [17, 18] it becomes apparent that indeed, there are great differences in the spread of these technologies. Nevertheless, the analyses of scientific literature place greater emphasis on two technological solutions: cloud technology and IoT-technology [19].

In relation to IoT-technology solutions, it is important to stress that these solutions are capable of the followings: They are capable of collecting data through sensors and making operational decisions when appropriate. The collected data can be forwarded data by connecting to a data network (LAN). This can take place in a centralized and/or decentralized manner known as M2M communication. And these solutions are capable of executing external decisions or decisions that were previously made autonomously in the form of control [20].

There have been a number of studies on the subject of digital transformation that address the difficulties of introducing Industry 4.0 technologies [16, 21]. Agostini-Nosella [22] studied the various Industry 3.0 technologies that can serve as a basis for Industry 4.0 solutions. By analogy, the study presented in my publication is based on the idea that the results of Industry 4.0 are technologically based on the IT solutions of Industry 3.0 [23, 24], connected to a somewhat integrated corporate governance system. The use of ICT infrastructure solutions tied to the Industry 3.0 "generation" is of great significance not merely because it is based on or supplements its solutions [10], but also because the driving force behind the implementation of Industry 4.0 solutions is that the management of for-profit organisations are capable of sensing and appreciating how ICT-solutions help companies stay afloat in the market competition [25]. The "source" of these challenges can lie within the economic operator (e.g. cost-efficiency) or can originate from outside the organization – for example, when the economic operator in question is an integral part of a larger, integrated supply chain. Amongst the various success factors of company IT systems, a number of authors mention the preparedness and attitude of market partners towards the implementation and integration of ICT solutions [26]. Of these, special mention must be made of the publication of Acar et al. [27], which used quantitative means to demonstrate the connection between the use of ERP, the integrating role of the supply chain and company efficiency. In his study, he followed a similar course [28] of measuring the company efficiency of respondents through "self-assessment" and did not rely on any financial-accounting data to measure the impact of ERP systems. An evaluation based on such a "self-report" can only be successful if workers using the information system share relevant information and experiences. This is also understood in the dimension of sharing knowledge and experience gained from the use of the information system, but also in the dimension of ensuring the flow of information within the company, so that top management can get a picture of the real benefits of using a particular ICT solution and the innovation potential that is to be exploited. The research by Mura et al. [29] shows precisely that this attitude is clearly evident in around 60% of companies, and is a problem in less than a fifth (19%) of companies.

The main aim of this research is to verify that successful ICT-innovation has certain external factors (such as the integration into a supply chain) and internal factors (e.g. the pursuit of efficiency) while it is also necessary for the decision-makers of for-profit organisations to sense the advantages of the ICT-

infrastructure [30]. This research is built on the assumption that the management of business organisations can make a sophisticated distinction between the exact nature of the benefits derived from ICT infrastructure. The main aim of this study to examine how typically "Industry 3.0" solutions of Hungarian SMEs set the stage in terms of technology and business for the regular use of Industry 4.0 solutions.

3 General Overview of the Survey

The results stated in the present publication were collected through a questionnaire-based survey conducted in two waves (from spring 2019 to autumn 2019). The questionnaire included a total of 78 questions and the purpose of the survey was to examine the ICT infrastructure and information security relations of the respondent for-profit organisations in the light of senior management satisfaction. The survey was conducted with non-anonymous means, which allowed for establishing connections between the responses and the data published in the respondent is financial reports. The questionnaire was structured in a way that allowed a senior manager to complete it in 12-15 minutes, accordingly placing emphasis on economic and IT-related questions.

Based on the financial reports submitted by respondents in 2018 pursuant to Hungarian accountancy rules, in my study the distribution of for-profit organisations involved in the survey were examined based on their balance sheet total and staff numbers. (Three of the respondents are only engaged in IFRS reporting, therefore, their information was not listed in the official databases could be used for the survey.) The classification of respondents based on balance sheet total and staff numbers are presented in Table 1 below:

Table 1
Distribution of respondents by balance sheet total and staff numbers (source: own ed.)

Staff	**Balance sheet total (in HUF)**					
	less than 1 million	**1–10 million**	**10-35 million**	**25-100 million**	**more than 100 million**	**Total**
> 10	90	4	0	0	0	94
11–50	211	41	0	0	0	252
50–250	30	87	5	3	1	126
250	1	10	10	2	0	23
Total	332	142	15	5	1	495

Left-sided asymmetry is clearly observable along both dimensions and presumably the survey results were greatly influenced by the significant overrepresentation of for-profit organisations with a low balance sheet total.

The respondents were not classified by industry as there was no available database to reliably classify the various businesses in one or more industries. The spot-check like analysis has led me to the conclusion that based merely on the scope of operations listed in the company register (TEÁOR number, unified sectoral classification system of economic activities), The respondents cannot be classified reliably in one or more industries. Nevertheless, for-profit organisations that are presumably required due to compliance with the regulatory environment to obligatorily use IT devices were not approached to complete the survey. The goal was to only have managers complete the survey who are the heads of for-profit organisations where the use of ICT devices is based solely on personal discretion.

4 Research Model and Analysis of the Results

In the sub-survey of the survey outlined in the previous chapter which is presented in the current publication, the respondents of the questionnaire are analyzed based on a simplified model of the transition from Industry 3.0 to Industry 4.0, while also examining the extent to which the senior management of the for-profit organisations are capable to sensing the advantages of the ICT infrastructure services. If this is true, then they are able to make accurate decisions in order to choose the right IT solution to achieve the competitive advantage they want to achieve. In this case, the company's management is able to perceive in time when and how to integrate Industry 4.0 solutions into the life of the company. The survey model is summarized in Figure 1 below:

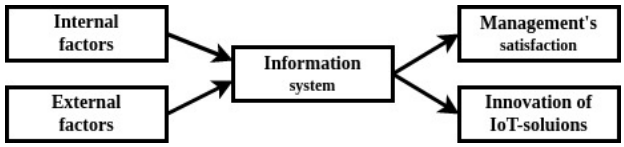


Figure 1
Theoretical model of the survey (source: own ed.)

Only the questions of the questionnaire will be processed for the verification of the model, and accounting data shall not be used for this sub-research.

4.1 Relevant Questions of the Questionnaire

In the following, the questions of the questionnaire and the manner of their processing will be presented. Hereinafter, the various questions shall be represented by their number and with variables representing the question (indicated in parentheses at the end of the question).

In the research presented in this publication, three groups of questions from the questionnaire were processed. Two out of the three groups of questions had to be answered in the same way. Question group 'A3' was used to assess the position of the company in the supply chain and to determine the character of the perceived competition in the market. On this onset, this set of questions asked about economic relations. Question group 'C1' asked about the company's use of classic IT infrastructure elements typical of the Industry 3.0 era and senior management satisfaction with the benefits of the ICT infrastructure. Questions belonging to these two groups had to be answered on a five-point Likert scale. This scale was in line with conventional Hungarian evaluations (1: the worst; 5: the best) and therefore the responses to the questions are on an ordinal scale. For the sake of standardized evaluation, for normalization, the transcoding was carried out as follows: answer number one was transcoded as 0, answer number two as 0.25, answer number three as 0.5, answer number four as 0.75 and answer number five as 1.

Question group 'D1' assessed the stage of adoption of IoT-based solutions. The respondents gave replies to the group of questions on a Likert scale of one too five. The five possible responses are as follows: “Do not use”, “Planned”, “Under implementation”, “Partly introduced” and “Introduced”. The possible responses greatly resemble the categories of Klisenko–Serral [31], although the research results were not available at the time the questionnaire was designed. The responses delineate a developmental trajectory. Therefore, the responses to the questions are on an ordinal scale. During transcoding, the responses were assigned values of 0, 0.25, 0.5, 0.75 and 1.

Each question is listed following the logic of Figure 1.

4.1.1 External Factors

External factors include the following questions:

A3/1. Our company is under considerable pressure to innovate (pinn).

A3/3. Procurements are made out electronically in an automated form (eprch).

A3/4. Sales are made out electronically in an automated form (esll).

Questions 3 and 4 are clearly questions that measure e-commerce activity. This becomes relevant because with the increasing intensity of e-commerce, it can be assumed that data exchange is increasingly automatic and that supply chain activity and related data exchange is increasingly auto-automated. The questionnaire was an attempt to distinguish between B2B and B2C activities. However, the variable representing this question showed no significant correlation to any other variables representing other questions, therefore, it was no longer a part of the analysis.

The innovative character of market competition has been considered an external factor because innovation pressures can be assumed to differentiate the products/services of each actor in the competitive market, so that significant operation influencing information is flowing from the market.

4.1.2 Internal Factors

Internal factors include the following questions:

A3/2. Our company is under considerable price competition pressure (pprc).

A3/6. The implementation/development of IT system(s) have had an impact on the company's structure and/or business processes (icteff).

Question 2 is included as an internal factor because in the case of price competition, it is assumed that there is a segment of the market where buyers are price sensitive, so that the path to market success is partly or entirely through lower prices. A lower price is achieved through lower costs, which implies significant internal efficiencies. The question 6 clearly asks about the impact of IT-Alligment, i.e. whether the respondent has gone through a process of harmonisation between the operation of the company and the operation of the ICT infrastructure.

4.1.3 Information System

The information system part consists of two questions. One question asks about the business processes that are the basis of the information systems, and the other examines the presence of integrated systems. An integrated system can be expected to support all or at least most of the business activities:

A3/5. Low (1) or high (5) degree of automation in production/services (arate).

C2/1. Using standalone software (1) or integrated (5) systems (intsys)?

4.1.4 Management's Satisfaction

The questions of the group C2 basically measure the satisfaction of the condition of the ICT-infrastructure and its operation. As the various questions of the questionnaire covered numerous fields, there are questions on the subjective experiences related to the operation of the ICT-infrastructure from five different approaches. Throughout the survey, respondents were asked about the efficiency-improving effect of the ICT-infrastructure on operative work (C2/2 and C2/4); its market position-strengthening effect (C3/3 and C2/5) as well as the ICT-infrastructure's strategic impact. The questions of the group C2 were the follows:

C2/1.We sense the benefits of the ICT-systems in efficient work (beffw).

C2/2. We sense the benefits of the ICT-systems in customer satisfaction (bcuss).

C2/3. We sense the benefits of the ICT-systems in automation of processes (bpaut).

C2/4. We sense the benefits of the ICT-systems in reacting to challanges (benrea).

C2/5. We sense the benefits of the ICT-systems in growth (bcuss).

5 Results and Findings

The results of the research are analysed in two steps: first, a descriptive statistical analysis of the responses is carried out (and this is complemented by an examination of the relationship between some of the variables), and then the model shown in Figure 1 is validated using regression models.

5.1 Descriptives

Supplementing the descriptive statistic review of the responses to the questions of the group A3, the Pearson's A skewness coefficient is summarized in the following table:

Table 2
Summary of the responses to the relevant questions of the Group A3 (source: own ed.)

Qn.	1	2	3	4	5	Avg.	Mode	Median	A-value
A3/1	74	94	146	111	73	3.03	3	3	0.02
A3/2	20	22	77	150	229	4.10	5	4	-1.85
A3/3	155	148	111	56	28	2.31	1	2	1.10
A3/4	331	84	49	22	12	1.59	1	1	0.60
A3/5	92	106	186	85	29	2.70	3	3	-0.26
A3/6	32	61	123	177	105	3.53	4	4	-0.42

The descriptive statistic data show that comparatively, the responses show widely ranging positional and mean values with a distribution of different degrees and directions. These factors gained relevance in the case of two pairs of questions, those concerning the extent of the digitalization of the supply chain (questions A3/3 and A3/4) as well as those referring to price competition and the market pressure to innovate (A3/1 and A3/2). The cross tabulates of the two questions referring to market competition took shape in Table 3.

The data show that unfortunately, price competition has a more intensive impact on the respondents, whilst price competition and innovation competition are non-exclusive as 30,72% of the respondents (n=153) were identified in significant price and innovation competition. Only 28.71% of the respondents (n=143) are on the main diagonal of Table 3.

Table 3
The impact of market and innovation competition on respondents (source: own ed.)

A3/2 (price competition)	A3/1 (innovation pressure)					
	1	**2**	**3**	**4**	**5**	**Total**
1	15	1	3	0	1	20
2	4	8	2	7	1	22
3	14	13	28	10	12	77
4	20	31	44	39	16	150
5	21	41	69	55	43	229
Total	74	94	146	111	73	498

This also implies that the vast majority of respondents (71.29%) are dominated by either price or quality-based competition that pushes innovation. This is also confirmed by the χ^2-test performed (χ^2=83,31, p<0,001). However, this allocation was not suitable for cluster analysis as there was no allocation with individual clusters that had roughly the same number of respondents. The examination of supply chain digitalization yielded similar results (Table 4):

Table 4
Supply chain digitalization amongst respondents (source: own ed.)

A3/4 (electronic sales)	A3/3 (electronic procurement)					
	1	**2**	**3**	**4**	**5**	**Total**
1	142	93	54	27	15	331
2	7	42	25	7	3	84
3	5	6	23	11	4	49
4	0	6	6	7	3	22
5	1	1	3	4	3	12
Total	155	148	111	56	28	498

Yet again, there are significant differences in terms of the number of respondents belonging to possible clusters and the relative predominance of electronic procurement over electronic sales is also considered unfortunate. When examining the absolute numbers, there are a significant number (n=281, i.e. 57%) of respondents who do not belong to some digital supply chain, or only to a minimum extent. The extent to the respondents' digital supply chain structures were analysed. Only 43.17% of respondents (n=215) are on the main graph in Table 4 (χ^2=119,21, p<0,001). Reviewing the distribution of the data in Table 4, we can see that electronic sales are more significant. Considering the demographic data, this can be interpreted as the fact that these are typically small and medium sized enterprises, which do not have a supplier base in which an integrated relationship could be established.

The descriptive statistics for question group C2 were as follows (Table 5):

Table 5
Summary of the responses to the relevant questions of the group C2 (source: own ed.)

Question	1	2	3	4	5	Avg	Mode	Median	A-value
C2/1	76	55	133	131	103	3.26	3	3	0.20
C2/2	35	34	104	173	152	3.75	4	4	-0.22
C2/3	45	69	138	132	114	3.40	3	3	0.33
C2/4	47	69	113	149	120	3.45	4	4	0.44
C2/5	43	76	127	153	99	3.38	4	4	-0.51
C2/6	52	83	137	143	83	3.24	4	3	-0.62

The review of satisfaction-related responses is worthy of descriptive statistic analysis. One of the essential features is that there is both noticeable right and left-sided asymmetry and there are significant differences in regard to positional and mean values. This is of great significance as this verifies that although multiple questions refer to the same area, these variations show that the questions examine different conditions. As the cross-tabulation of satisfaction-related questions would require 10 cross-tabulates, a correlation matrix in the Table 6 for further examination has been employed:

Table 6
The correlation matrix of satisfaction-related questions (source: own ed.)

	C2/2	C2/3	C2/4	C2/5	C2/6
C2/2	1.00	0.77***	0.76***	0.67***	0.64***
C2/3		1.00	0.75***	0.77***	0.72***
C2/4			1.00	0.77***	0.76***
C2/5				1.00	0.81***
C2/6					1.00

In the correlation table above, the significance level for the correlation of all pairs of variables is $p<0.001$ (***). The correlation matrix data confirms the descriptive statistics data, i.e. the responses to the questions are significantly interrelated, yet the scale of difference is enough to ensure they serve as the subject of sophisticated analysis. (This means that the respondents did not evaluate the same phenomenon in their responses given to the various questions.)

The descriptive statistics for question group D1 were as follows:

Table 7
Summary of the responses to the relevant questions of the group A3 of questions (source: own ed.)

Qn.	0.00	0.25	0.5	0.75	1.00	Avg.	Mode	Median	A-value
D1/1	191	77	30	168	32	2.54	1	2	1.07
D1/2	262	56	25	119	36	2.22	1	1	0.83
D1/3	208	81	30	103	76	2.51	1	2	0.97
D1/4	298	50	28	96	26	2.00	1	1	0.73

A number of phenomena are to be emphasized in the statistical analysis. One of these is that the responses to the four examined questions show a distribution with similar "characteristics" (with right-skewed multimodal empirical distribution), yet there are still significant variations and differences in regard to positional and mean values. Of these, the two questions must be highlighted on IoT infrastructure of which the crosstabulate is as follows (Table 8):

Table 8

The use of IoT technologies amongst respondents (source: own ed.)

D1/2. we are capable of the remote control of production devices (IoT).	**D1/1. we are capable of automated data collection (IoT)**					
	0	**0.25**	**0.5**	**0.75**	**1**	**Total**
0	161	30	5	59	7	262
0.25	7	25	10	12	2	56
0.5	2	4	8	10	1	25
0.75	12	14	6	75	12	119
1	9	4	1	12	10	36
Total	191	77	30	168	32	498

In relation to the table, it must be noted that only 279 of the 498 respondents (56%) are in the "major diagonal", i.e. automated data collection and control have similar importance in their production. This could be due to the use of autonomous vehicles and devices [1, 32], yet it is definitely noteworthy that there are 119 responses indicating the two extremes, i.e. when a company solely uses IoT solutions for either data collection or control.

The fact that all the responses for the three groups of questions form an ordinal scale means that the connection between the sub-questions of the three questions can be analyzed with inferential statistics.

4.2 Model Creation Results

Throughout the model creation process it became clear that there are two questions that may play a central role in the model (Figure 1: "Information System" element). These are question A3/5 on the automation of business processes and question C2/1 on the application of integration systems. This approach is in line with the [33, 34], pointing out the indispensable role of business processes (amongst others) in terms of the function of the ICT-infrastructure, the execution of which is realized by certain elements of the ICT-infrastructures of for-profit organisations. [1, 3, 4] When examining the relevant lines of tables 1 and 4, it is noticeable that they show an empirical distribution with different kurtosis, yet identical size – close to zero. These two elements correspond to the central element of pre-Industry 4.0 generation information

systems [6]. (Naturally, the stochastic relation between these two factors has been examined, yet the explanatory power of the established regression model was so low – R^2=0.044 – that the strength of the connection between the two can be negligible.) The external and internal factors of the model will be grouped around these two external elements. In regard to the exploration of the factors explaining the two central elements, the model creation was concluded with the following results:

$$\text{arate} = 0.2482^{***} + 0.2269*\text{pinn}^{***} + 0.0963*\text{eprch}^{*} + 0.2100*\text{esll}^{***} \quad (1)$$

where the global testing confirmed the significance of the model (F3, 494 = 28.35, p < 2.2e-16; adjusted R^2 = 0.1417, significance of variables: ***: p < 0.001; *: p < 0.05.)

$$\text{intsys} = 0.2654^{***} + 0.0974*\text{pprc}^{\circ} + 0.3553*\text{icteff}^{***} \quad (2)$$

where the global testing confirmed the significance of the model ($F_{2, 495}$ = 29.24, p < 0,001; adjusted R^2 = 0.1057, significance of variables: ***: p < 0.001; °: p < 0.1.)

Throughout the model creation process, in relation to both outcome variables, It was examined, whether they are significantly related to any of the possible explanatory variables. In the case of equations (1) and (2), only significant explanatory variables were displayed. Equations (1) and (2) are summarized in the following Figure 2:

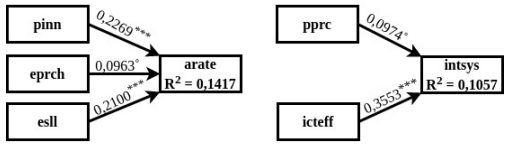


Figure 2
Equations (1) and (2) in a unified model (source: own ed.)

Since the relationship between the information system and management satisfaction would be described by five regression models (Figure 3), the actual models will be forgo described and will instead present a summary of the results similarly to Figure 2 (8 additional regression equations would have to be described, which would not yield any additional information beyond the summary figure, yet their description would presumably hinder the clarity of the overview, therefore these regression models will not be discussed.):

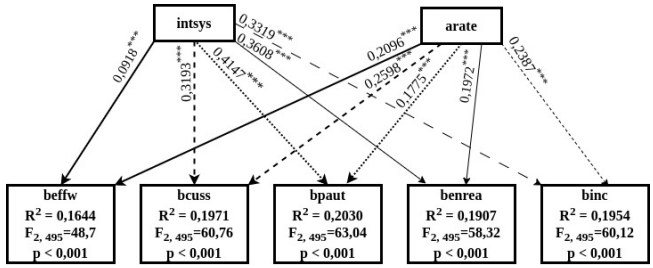


Figure 3

Sophisticated recognition of the ICT-infrastructure's benefits by the management (source: own ed.)

The analysis will proceed in Figure 4 in the same fashion as in the case of the analysis of the relationship between the "Information System" representing Industry 3.0 and the basic technologies of Industry 4.0.

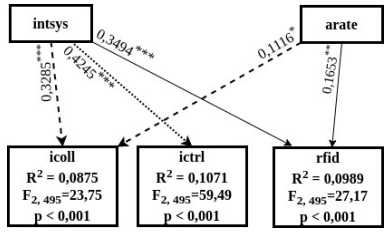


Figure 4

Relationship between the existing ICT-Infrastructure and Industry 4.0 (source: own ed.)

In the case of Figures no. 5, 6 and 7, the significance level is indicated as usual: ***: $p < 0.001$; **: $p < 0.001$; *: $p < 0.05$°·: $p < 0.1$.

Figures no. 5, 6 and 7 show that a significant statistical relationship was established successfully between external and internal factors and the ICT-infrastructure (more specifically, the integrated corporate governance system and business processes), as well as between the ICT-infrastructure and senior management satisfaction and the basic technologies of Industry 4.0. Therefore, the verification of the survey's theoretical model was a success. The qualitative evaluation of the results can take place accordingly.

6 Evaluation of Results

As first step Figure no. 5 will be evaluated, which aims to provide a summary of the entire model. It is important to emphasize that the degree of the automation of business processes and the existence of an integrated corporate governance system are the result of the combined effect of the company's external and internal

factors. The question is whether the integrated corporate governance system had a significant explanatory power on the entirety of the company. Consequently, the partial conclusion can be drawn that the introduction of integrated systems is a precondition for the implementation of Industry 4.0 technologies, not just in terms of technology, but also this requires harmonizing the IT and business processes within the economic operator.

Apparently, the role the company plays in a supply chain and the market competition it comes up against also have an identifiable and statistically verifiable effect on the company ICT-infrastructure (the extent of innovation and price competition pressure).

Figure no. 6 provided a summary illustration of the fact that the automation of business processes and the existence of integrated systems are necessary for company managers to sense the advantages of the ICT-infrastructure. Although the regression models of the summary Figure have a similarly explanatory power, the coefficients of the explanatory variables in the five regression models – where the explanatory variables were identical and only the result variables were different –diverge significantly. It follows that in themselves, the two elements contribute in different degrees to the tangible "success" of the ICT-environment. Therefore, the various elements of the ICT-infrastructure can have their full effect in synergy with other (hard and soft) organizational factors. However, throughout the examination of Industry 4.0 solutions, this also suggests that it is worth considering the question of the other supporting factors required for the realization of competition advantages.

The three regression models summarized in Figure no. 7 verify the generational connection between Industry 3.0 and Industry 4.0. In line with previous experiences, the three regression models have a similar explanatory power. Oddly enough, there is no statistically significant relationship between automatic control devices and integrated systems. Two consequences can be drawn from this: firstly, it is not a worthwhile effort to uniformly assess the impact of IoT devices on for-profit organisations in the sense that not all devices capable of automated data collection have the same impact as devices that are capable of functioning remotely. Secondly, when it comes to IoT devices, remotely and/or centrally controlled devices and autonomous production devices and vehicles should be assessed separately in terms of control [6, 8]. Generally speaking, there are two different types of IoT-infrastructures that can and should be distinguished and evaluated separately: the infrastructure of centralized and decentralized IoT-devices should be examined side by side.

Returning to Figure no. 1: throughout the statistical analyses the statistically significant relationship between the senior management satisfaction of the use of the ICT-infrastructure and the implementation of Industry 4.0 devices will be demonstrated, yet this ultimately proved to be unsuccessful. However, the presented interim conclusions are sufficient for drawing up a Technology

Acceptance Model [35] – with the inclusion of further variables representing other questions included in the questionnaire – with the aim of identifying the factors that support or hinder the IT innovations of Hungarian small and mid-size enterprises.

Conclusions

Two conclusions can be drawn from the survey results. Firstly, in themselves, Industry 4.0 solutions cannot ensure any competitive advantages. For this to truly represent a competitive advantage, it is necessary to understand the external surroundings of the company – even if these are not market-driving factors due to the size and/or significance of the company. Thus, the innovation of the ICT-infrastructure must suit the internal environment of the company and must also take the market environment into consideration. As it from the survey results can be seen, the management of for-profit organisations are capable of the sophisticated evaluation of the advantages offered by the ICT-infrastructure, yet this in turn implies that one of the preconditions of successful ICT-investments is that the management of for-profit organisations must be aware of the driving forces behind the market competition and consequently, be able to identify in which field and to what extent they require the added value of ICT-infrastructure services.

As mentioned above, the spread of Industry 4.0 solutions is revolutionizing the production processes of industries that were previously considered to be non-technology intensive. [8, 11] The interim conclusions of my survey presented in this publication indicate – as a warning sign for the future – that the lack of inclusion of a particular technological generation can lead to great subsequent strategic disadvantages as it leads to the absence of the technological environment that can serve as the basis for the appearance of a new generation of technology. However, it remains to be seen what impact the advent of revolutionary changes necessitated by "deferred" evolutionary changes shall have on the economic operator in question.

References

[1] C. Bai, P. Dallasega, G. Orzes and J. Sarkis, "Industry 4.0 technologies assessment: a sustainability perspective," *International Journal of Production Economics*, Vol. 229, Nov. 2020, Art. no. 107776, doi: 10.1016/j.ijpe.2020.107776

[2] J. L. Hopkins, "An investigation into emerging industry 4.0 technologies as drivers of supply chain innovation in Australia," *Computers in Industry*, Vol. 125, Feb. 2021, Art. no. 103323, doi: 10.1016/j.compind.2020.103323

[3] A. Keszthelyi, "Some Special Results of ICT Revolution," in *XVIII. International May Conference on Strategic Management*, in IMCSM Proceedings, Vol. 18, No. 1, May 2022, pp. 479-484

[4] H. Ç. Bal and C. Erkan, "Industry 4.0 and Competitiveness," *Procedia Computer Science,* Vol. 158, pp. 625-631, 2019, doi: 10.1016/j.procs.2019.09.09

[5] M. E. Porter, "The Value Chain and Competitive Advantage," in *Understanding Business: Processes*, D. Barnes, D, Ed., London, United Kingdom: Routledge, 2001, pp. 50-66

[6] S. Aheleroff, X. Xu, Y. Lu, M. Aristizabal, J. P. Velásquez, B. Joa and Y. Valencia, "IoT-enabled smart appliances under industry 4.0: A case study," *Advanced Engineering Informatics*, Vol. 43, Jan. 2020, Art. no. 101043, doi: 10.1016/j.aei.2020.101043

[7] P. Pisar and D. Bilkova, "Controlling as a tool for SME management with an emphasis on innovations in the context of Industry 4.0," *Equilibrium. Quarterly Journal of Economics and Economic Policy*, Vol. 14, No. 4, pp. 763-785, Dec. 2019, doi: 10.24136/eq.2019.035

[8] M. Alexy and T. Haidegger, "Precision Solutions in Livestock Farming– feasibility and applicability of digital data collection," in *2022 IEEE 10th Jubilee International Conference on Computational Cybernetics and Cyber-Medical Systems (ICCC)*, 2022, pp. 233-238, doi: 10.1109/ICCC202255925.2022.9922883

[9] F. Othmane, M. A. Ferrag, L. Shu, L. Maglaras and X. Wang, "Internet of Things for the Future of Smart Agriculture: A Comprehensive Survey of Emerging Technologies," *IEEE/CAA Journal of Automatica Sinica*, Vol. 8, No. 4, pp. 718-752, Apr. 2021, doi: 10.1109/JAS.2021.1003925

[10] D. A. Zakoldaev, A. V. Shukalov, I. O. Zharinov and O. O. Zharinov, "Modernization stages of the Industry 3.0 company and projection route for the Industry 4.0 virtual factory," in *IOP Conference Series: Materials Science and Engineering,* Vol. 537, No. 7, 2019, Art. no. 032005, doi: 10.1088/1757-899X/537/3/032005

[11] M. Dobler and J. Schumacher, "Innovation and emerging ICT in logistics for Europe position paper for the workshop is logistics catching up the pace of emerging technologies?" in *International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, R. Jardim-Gonçalves et al., Eds, 2017, pp. 1509-1516, doi: 10.1109/ICE.2017.8280061

[12] N. N. Misra, Y. Dixit, A. Al-Mallahi, M. S. Bhullar, R. Upadhyay and A. Martynenko, "IoT, Big Data, and Artificial Intelligence in Agriculture and Food Industry," *IEEE Internet of Things Journal*, Vol. 9, No. 9, pp. 6305-6324, May 2022, doi: 10.1109/JIOT.2020.2998584

[13] J. Kárpáti-Daróczi, A. Tick and R. Saáry, "Implementation of Industry 4.0 among SMEs in Hungary" in: *Az üzleti szervezetek túlélési esélyei napjaink legújabb kihívásainak idején: Vállalkozásfejlesztés a XXI. században*, J. Varga et al., Ed. (eds.), Vol. 2022/1, 2022, pp. 141-160

[14] A. Tick, "Industry 4.0 Narratives through the Eyes of SMEs in V4 Countries, Serbia and Bulgaria," *Acta Polytechnica Hungarica*, Vol. 20, No. 2, pp. 83-104, 2023, doi: 10.12700/APH.20.2.2023.2.5

[15] A. Tick, J. Kárpáti-Daróczi, and R. Saáry, "To familiarise or not to familiarise' - industry 4.0 implementation in SMEs in Hungary," in *Possibilities and barriers for Industry 4.0 implementation in SMEs in V4 countries and Serbia*, Ed., Bor, Serbia: University of Belgrade, Technical Faculty in Bor, Engineering Management Department, 2022, pp. 35-61

[16] T. Masood and P. Sonntag, "Industry 4.0: Adoption challenges and benefits for SMEs," *Computers in Industry*, Vol. 121, Oct. 2020, Art. no. 103261, doi: 10.1016/j.compind.2020.103261

[17] J. Basl, "Enterprise Information Systems and Technologies in Czech Companies from the Perspective of Trends in Industry 4.0," in *Research and Practical Issues of Enterprise Information Systems. CONFENIS 2016.*, in Lecture Notes in Business Information Processing, Vol 268. Nov. 2016, pp. 156-165, doi: 10.1007/978-3-319-49944-4_12

[18] Zs. R. Szabó and L. Hortoványi, "Digital transformation and Industry 4.0: experiences from Hungary, Serbia, Slovakia and Romania," *Külgazdaság,* Vol. 65, Nos. 5-6, pp. 56-76, Jun. 2021, doi: 10.47630/KULG.2021.65.5-6.56

[19] B. Chen, J. Wan, L. Shu., P. Li, M. Mukherjee and B. Yin, "Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges," IEEE Access, Vol. 6, pp. 6505-6519, 2018, doi: 10.1109/ACCESS.2017.2783682

[20] V. Majstorovic, S. Stojadinovic, B. Lalic and U. Marjanovic, "ERP in Industry 4.0 Context," in *Advances in Production Management Systems. The Path to Digital Transformation and Innovation of Production Management Systems. APMS 2020.* in IFIP Advances in Information and Communication Technology, Vol. 591, 2020, doi: 10.1007/978-3-030-57993-7_33

[21] G. Orzes, E. Rauch, S. Bednar and R. Poklemba, "Industry 4.0 Implementation Barriers in Small and Medium Sized Enterprises: A Focus Group Study," in 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2018, pp. 1348-1352, doi: 10.1109/IEEM.2018.8607477

[22] L. Agostini and A. Nosella, "The adoption of Industry 4.0 technologies in SMEs: results of an international study," *Management Decision,* Vol. 58, No. 4, pp. 625-643, 2020, doi: 10.1108/MD-09-2018-0973

[23] Z. Jiang, S. Yuan, J. Ma and Q. Wang, "The evolution of production scheduling from Industry 3.0 through Industry 4.0.", *International Journal*

of Production Research, Vol. 60, No. 11, pp. 3534-3554, 2021, doi: 10.1080/00207543.2021.1925772

[24] Y. Yin, K. E. Stecke and D. Li, "The evolution of production systems from Industry 2.0 through Industry 4.0," *International Journal of Production Research*, Vol. 56, No. 1-2, pp. 848-861, 2017, doi: 10.1080/00207543.2017.1403664

[25] T. Taurino and A. Villa, "A method for applying Industry 4.0 in Small Enterprises," *IFAC-PapersOnLine*, Vol. 52, No. 13, pp. 439-444, 2019, doi: 10.1016/j.ifacol.2019.11.099

[26] H. O Awa and O. U. Ojiabo, "A model of adoption determinants of ERP within T-O-E framework," *Information Technology & People,* Vol. 29, No. 4, pp. 901-930, 2016, doi: 10.1108/itp-03-2015-0068

[27] M. F. Acar, S. Zaim, M. Isik and F. Calisir, "Relationships among ERP, supply chain orientation and operational performance," *Benchmarking: An International Journal,* Vol. 24 No. 5, pp. 1291-1308, 2017, doi: 10.1108/bij-11-2015-0116

[28] D. Gefen, A. Ragowsky, "A Multi-Level Approach to Measuring the Benefits of an Erp System in Manufacturing Firms," *Information Systems Management*, Vol. 22, No.1, pp. 18-25, 2005, doi: 10.1201/1078/44912.22.1.20051201/85735.3

[29] L. Mura, T. Zsigmond, and R. Machová, "The effects of emotional intelligence and ethics of SME employees on knowledge sharing in Central-European countries," Oeconomia Copernicana, Vol. 12, No. 4, pp. 907-934, 2021, doi: 10.24136/oc.2021.030

[30] D. Horváth, R. Zs. Szabó, "Driving forces and barriers of Industry 4.0: Do multinational and small and medium-sized companies have equal opportunities?" *Technological Forecasting and Social Change*, Vol. 146, pp. 119-132, 2019, doi: 10.1016/j.techfore.2019.05.021

[31] O. Klisenko and A. E. Serral, "Towards a Maturity Model for IoT Adoption by B2C Companies," *Applied Sciences,* Vol. 12, No. 3, 2022, Art. no. 982, doi: 10.3390/app12030982

[32] F. Othmane, M. A. Ferrag, L. Shu, L. Maglaras and Wang, X., "Internet of Things for the Future of Smart Agriculture: A Comprehensive Survey of Emerging Technologies," *IEEE/CAA Journal of Automatica Sinica*, Vol. 8, No. 4, pp. 718-752, Apr. 2021, doi: 10.1109/JAS.2021.1003925

[33] P. Michelberger, S. Dombora, "A Possible Tool for Development of Information Security-Siem System," *Ekonomika, Journal for Economic Theory and Practice and Social Issues*, Vol. 62, No. 1, Mar. 2016

[34] P. Michelberger, Zs. Horváth, "Security aspects of process resource planning," *Polish Journal of Management Studies*, Vol. 16, No. 1, pp. 142-153, 2017, doi: 10.17512/pjms.2017.16.1.12

[35] Folkinshteyn, D. and Lennon, M., „Braving Bitcoin: A technology acceptance model (TAM) analysis," *Journal of Information Technology Case and Application Research*, Vol. 18, No. 4, pp. 220-249, Feb. 2016, doi:10.1080/15228053.2016.1275242

A Hybrid Machine Learning-based Control Strategy for Autonomous Driving Optimization

Ahmad Reda, Rabab Benotsmane, Ahmed Bouzid, József Vásárhelyi

Institute of Automation and Info-communication, University of Miskolc, Egyetemváros, 3515 Miskolc, Hungary, {autareda, iitrabab, ahmed.bouzid, vajo}@uni-miskolc.hu

Abstract: Developing autonomous vehicles is a highly important topic in the field of intelligent transportation systems. Automated steering is a crucial function in the autonomous vehicle. Therefore, it is urgent to either develop a new effective control strategy or improve existing ones. A variety of control strategies are used for this purpose, most with limitations related to their computing capabilities with the highly complex systems or to lack of efficacy related to maintaining the balance between driving performance and driving smoothness. In this paper, three different machine learning-based models were developed to perform an autonomous driving task: a supervised learning model (Deep Neural Network, DNN), a reinforcement Deep Q-learning model (DQN), and a hybrid model. The DNN model was trained based on the behavior of the classical MPC controller. The DQN was designed with the same structure as the DNN and trained by directly interacting with the driving environment. The hybrid model is a combination of supervised and reinforcement learning algorithms, where the trained DNN model is used as a decision-maker (Actor) in a deep deterministic policy gradient reinforcement learning model. The behavior of the designed models was compared based on several performance indicators, including the ability to drive the vehicle along the desired trajectory, the response time, and the smoothness of the driving system. The results show that the DNN model was able to imitate the behavior of the traditional MP Controller efficiently and all three machine learning models successfully drive the vehicle along the desired path. The hybrid model achieves the best results and improved the smoothness of the driving system with a reasonable response time.

Keywords: Autonomous Driving; Model Predictive Control (MPC); Supervised Learning; Deep Neural Networks; Reinforcement Learning; Deep Q-Network (DQN); Deep Deterministic Policy Gradients (DDPG)

1 Introduction

The evolution of autonomous driving systems has seen the use of different technologies aiming to improve efficiency, enhance driving safety and reduce the risks related to traffic congestion. Driving in a structured environment and highway

driving projects were some of the earliest autonomous vehicle projects, carried out at Carnegie Mellon University and Bundeswehr University Munich [2], [3]. Since then, projects and research related to autonomous vehicles have been carried out by academic institutes and companies alike. According to the Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems "SAE-J3016", vehicle autonomy is divided into six different levels. Level 0 (No Automation) depends on the human driver to perform all the driving tasks, it is manually controlled. Level 1 (Driver Assistance) is considered the lowest automation level, where the driver has full responsibility, but some assistant driving systems are included for certain circumstances. Level 2 (Partial Automation) combines different automated functions which can be working simultaneously, such as steering and acceleration tasks, but the driver is still involved in the driving tasks such as performing the maneuvers and has to monitor the environment all the time. At Level 3 (Conditional Automation) the vehicle has the capability of detecting the surrounding environment and making decisions in normal conditions, but the necessity of the driver still exists, meaning that the driver has to be ready to take control over the vehicle at any time. At Level 4 (High Automation) the vehicle performs all the driving tasks in most circumstances, and the driver still has the option to take control. At Level 5 (Full Automation) the vehicle is capable of performing all driving tasks in all circumstances, and the driver has the option to manually override [4], [5]. The vehicle interacts with the surrounding environment in order to perform several related tasks: perception, where the required information about the driving environment is provided to the system; planning, where the optimal scenarios and the control actions are obtained based on the provided information; and the control function, where the control strategy is put into action [6]. The automated steering task is a part of the control function, where the tracking errors are minimized in order to follow the desired trajectory. Driving the vehicle along the desired trajectory is considered one of the most critical tasks due to the fact that any failure in the applied control strategy can have severe consequences. A variety of control strategies have been used to perform the automated steering task, such as the classical feedback control algorithm, Model-Based Control, Dynamic Control, and Adaptive Control [7], [8], [9], [10]. In this context, Model Predictive Control (MPC) has become the most commonly used algorithm for the autonomous vehicle steering system. The MPC controller solves an online optimization problem with the ability to handle the system constraints (soft-hard) by including them in the design process, which makes it a powerful strategy to deal with the stability and the changing dynamics of the vehicle. On the other hand, with the increase of the system complexity, the computational load of the MPC controller is increased, since it solves the optimization problem in each time step, and it may not be able to meet the real-time requirements. Additionally, MPC is resource-consuming, which makes it invisible, especially when it comes to the limited resources of embedded computing platforms such as system-on-chip (SoC) and field-programmable gate array (FPGA) adaptive platforms [11], [12]. Recently, Deep Neural Network (DNN) has gained attention and has been rapidly developed

and efficiently implemented with a variety of applications in different fields such as image classification [14], natural language processing, and speech recognition [15]. In contrast, to the classical control algorithms, which are mainly based on tuning predefined parameters related to a determined environment [16], the behavior of the deep neural network model is optimized based on the provided information (self-optimized algorithm). In other words, the neural networks algorithm bypasses the need for significant parameter tuning, which makes it more efficient to model highly complex systems and to deal with unforeseen situations, especially after being well trained and validated using sufficient datasets. Recently the implementation of deep neural networks within the domain of robotic applications has made massive progress and has provided promising results such as perception and motion planning [17] and object detection and semantic segmentation [18]. In contrast, to supervised learning, agents in Reinforcement Learning (RL) are trained by directly interacting with their environment rather than explicitly guiding the model on how to act based on the labeled data [19]. The performance of the RL agent is evaluated based on the reward function, where the agent is trained to act in the environment in a way that maximizes the cumulative reward in order to improve the performance [20]. RL has proven to be a powerful method mainly in the domains of game playing and robotic manipulation [21], [22], and RL algorithms are considered a promising potential solution for many other applications, especially in cases where classical supervised learning is not applicable. Although there are promising results achieved by the implementations of reinforcement learning with different complex tasks related to automated driving, RL is still an emergent field in this domain, where the implementations and deployment of real-world applications are still very much an open challenge and RL has not yet been applied to practice as successfully as supervised and unsupervised learning. The main contributions of this work can be summarized in two main points. The first is leveraging the advantages of reinforcement learning and supervised learning by combining them in one control model in such a way that the RL-based network optimizes the action that is taken by the supervised neural network (DNN) and achieves a better generalization capability with the complex driving environment. The second contribution comes in enriching the research on RL algorithms and paving the way to bring RL closer to real-world implementations. In [13], a classic MPC controller was designed and deployed on FPGA for automated driving task, while in this paper three different machine learning-based models are developed for the same task and compared to the traditional MPC. The first model is a DNN-based model, which is designed and trained using a supervised dataset obtained from the behavior of the classical MPC controller. The second model is a reinforcement learning-based model (DQN) which is designed and trained without any supervision data, but directly by interaction with the environment. The third model is a hybrid one, which is a combination between the DNN and reinforcement learning methods. The trained DNN will be used as decision maker working beside another network (critic) within a DDPG reinforcement model. The combined method is expected to provide an

optimized solution, as the actions that are taken by the decision maker (trained DNN) will be evaluated and optimized by another neural network in order to minimize errors. Additionally, the combined model will be able to deal with and adapt to new cases that have not been faced during training.

The paper is organized and structured as follows: The second section provides background, including the most common vehicle models and control strategies that are used for autonomous driving, in addition to the work related machine learning algorithms describing the main features and their implementations in the field of autonomous driving. In the third section, the MPC controller and the design of the suggested models are discussed. The implementations and the obtained results are analyzed and discussed in the fourth section. Finally, the conclusions are provided in the last section.

2 Background

In this section, an overview of the vehicle models, the control strategies of the path tracking task, and the related machine learning algorithms are described.

2.1 Path Tracking and Related Works

Path tracking can be categorized into three main groups: geometric, kinematic, and dynamic. Due to its simplicity, geometric path tracking is one of the most commonly used models. In a geometric vehicle model, only the dimensions and the position of the vehicle are taken into consideration with no regard to internal or external forces, velocity, or acceleration. Geometric controllers are the most common controllers in the field of path tracking due to their stability and simplicity, where the state variables are simple with the absence of the derivatives. Follow the Carrot, Pure Pursuit, and Stanley are the best-known geometric control strategies [23]. Unlike the geometric vehicle model, the kinematic model describes the motion of the vehicle taking into consideration the velocity and the acceleration with no regard to its internal forces [24], [25]. Several interesting studies have emerged in regard to kinematic controlling. Sun et al. [26] presented a study to address the problem of path tracking for the autonomous vehicle and analyze the relationship between the road model and path tracking method. De Luca et al. [27] provided a comparison study of different feedback solutions for different tasks such as path tracking and stabilization for a car-like robot (kinematic model). Kinematic and Geometric models are effective for systems where there is no need to take the internal and external forces into consideration. However, these forces should be taken into consideration under specific conditions such as a sharp trajectory curvature. Ignoring the vehicle dynamics under such conditions will negatively affect the performance and the safety aspects. In a dynamic model, the motion of

the vehicle is described with respect to its position, velocity and □ccelerateon, taking into considerations the applied internal and external forces such as the gravity force [28], [29]. Taking the effects of the vehicle dynamics into consideration naturally makes the dynamic controllers more efficient and stable than geometric and kinematic controllers [30]. However, dynamic feedback (such as the torque) is required for these control strategies, which in turns requires special types of sensors and more data processing. Consequently, dynamic controllers are more expensive in terms of the cost and computational loads [31]. An adaptive controller is also used for autonomous vehicle tasks, developed to deal with systems which have uncertain, unknown, or changeable parameters. Martins et al. [32] used an adaptive controller for a vehicle path tracking task and their proposed model used the linear and angular velocity as a reference signal. Artificial intelligence is widely used with adaptive controllers in order to improve the control decisions in terms of speed and accuracy. In paper [33], a lateral motion control method was provided where the objective of the suggested method is to maintain the yaw stability and minimize the tracking error. The control schema consists of two main modules, a steering controller to ensure the yaw stability and an artificial neural network approximator to estimate cornering stiffness uncertainty. In the field of AI in learning and control, many related works are highlighted dealing with linear and nonlinear controllers. In [34], [35], the authors of both papers use the linear controller as the classical PID and Fuzzy controller for a linear system [36], [37], while others have focused on using nonlinear controllers and learning algorithms as presented in [38], [39], [40]. Reference [41] reports a new Reinforcement Learning (RL)-based control approach that uses Policy Iteration (PI) and a metaheuristic Grey Wolf Optimizer (GWO) algorithm to train the Neural Networks (NNs). The GWO algorithm shows good results in NN training and solving complex optimization problems.

2.2 Reinforcement Learning Algorithms and Related Works

Sequential decision making problems can be formulated by Markov Decision Processes (MDPs), which is considered a bedrock of the problems that reinforcement learning solves. MDPs consist of a decision maker (agent), set of states (S), set of actions (T), transaction function (A), and reward function (R,) which can be represented as a tuple <S, A, T, R>. At each time step (t), and based on the received state ($S_t \in S$), the agent takes an action ($A_t \in A$) which represents a pair (A_t, S_t) in the next time step. Based on the taken action the environment is transitioned to a new $S_{t+1} \in S$, and the agent receives a reward $R_{t+1} \in R$, [42], [43] (see Figure 1). The cumulative reward is simply represented as a sum of the expected return at each time step. The probability of selecting an action by the agent from all possible actions at all possible states is determined by the policy (π) that the agent follows. In addition to the probability of the selection action, the value function evaluates how good it is for the agent to select an action at a given state under a policy (π), and this is called the action-value function ($q\pi$), or how good

it is for the agent to be at a given state following a policy (π), and this is called the state-value function (v_π). Equations 1 and 2 are the mathematical representations of the action-value and the state-value functions, respectively. The action-value function $q_\pi(s,a)$ is the expected reward ($\sum_{k=1}^{\infty}\gamma^k R_{t+k+1}$) starting from state (s) at time (t), performing the action (a) and following the policy (π), where the state-value function v_π(s) is the expected reward starting from state (s) at time (t) and following the policy (π). It is worth mentioning that q_π is also referred to as the Q-function and its output is called the Q-value (the quality of taking an action). In terms of optimality, the main goal of the RL algorithm is to select the optimal policy that will yield the highest expected reward for each state. The optimal policy is associated with an optimal state-value function (v_*) and an optimal action-value function (q_*) or optimal Q-function, which are represented in equations 3 and 4, respectively. The fundamental property that the optimal Q-function (q_*) must satisfy is the Bellman equation (see equation 5), where (R_{t+1}) is the expected reward that the agent obtains by taking the action (a) at state (s), whereas $\gamma \max q_*(s',a')$ is the maximum expected discounted reward that can be received from any next state-action pair [44], [45]. Reinforcement learning is a category of machine learning that studies the behavior of an agent and focuses on how this agent might interact with its environment. The main goal of the agent is to maximize the cumulative given rewards it receives over time in order to optimize its behavior in such an environment [46]. Based on the fact that the agent is able to learn the value function estimates or/and the policies directly, RL methods can be categorized into three main methods: value-based methods, policy-based methods, and actor-critical methods [47]. All of the methods share the same strategy of determining the actions and evaluating the agent behavior, but the essential difference is where the optimality resides.

$$q_\pi(s,a) = E_\pi\left(\sum_{k=1}^{\infty}\gamma^k R_{t+k+1} \,|S_t = s,\ A_t = a\right) \tag{1}$$

$$v_\pi(s) = E_\pi\left(\sum_{k=1}^{\infty}\gamma^k R_{t+k+1} \,|S_t = s\right) \tag{2}$$

$$v_*(s) = \max_\pi v_\pi(s) \tag{3}$$

$$q_*(s,a) = \max_\pi q_\pi(s) \tag{4}$$

$$q_*(s,a) = E_\pi\left(R_{t+1} + \gamma \max_{a'} q_*(s',a')\right) \tag{5}$$

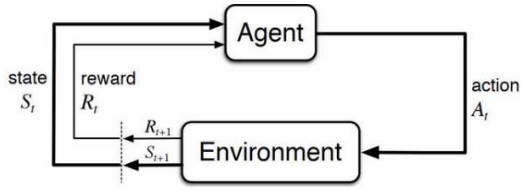


Figure 1
Markov Decision Processes

2.2.1 Value-based Algorithm

Value-based methods aim to get the optimal cumulative reward and determine the optimal policy that follows the recommendations. One of the most commonly used reinforcement learning value-based algorithms is the Q-learning method [48]. The objective of Q-learning is to find the optimal policy by learning how to find the optimal Q-value for the (s, a) pair, where the Q-values are stored in a Q-table. The Q-learning algorithm uses what is called the value iteration approach to converge the Q-function to the optimal Q-function by iteratively updating the Q-value for each (s, a) pair using the Bellman equation. With the increase in environment complexity, the state space size increases, and the performance of the Q-learning method will drop off because of the value iteration strategy that is used to update the Q-values (Q-table). The problem with large MDPs is that there are too many states and/or actions to be stored in the memory, and it is too slow to calculate the value for every individual state [49]. To overcome this problem, a function approximation is used to estimate the values instead of using the value iteration. The deep neural network is used as a function approximation and combined with the Q-learning method. This method is called Deep Q-learning, where the Deep Q-Network (DQN) approximates the Optimal-Q value [50]. The DQN model accepts the state as an input and outputs the estimated Q-value for every possible action that can be taken at that given state. After calculating the loss, the weights within the neural network are updated by stochastic gradient descent (SGD), just like in any other neural network.

2.2.2 Policy-based Algorithm

Like the value-based method, the policy-based method selects one possible action and evaluates the agent's behavior thereafter in order to achieve optimization. The essential difference between the two methods is a matter of how to achieve optimality. While the value-based method selects the optimal policy based on the optimal cumulative reward, the policy-based method directly optimizes the policy itself. The policy is parameterizes $\pi_\theta(s, a)$ and the optimization problem turns out to be finding θ, which maximizes the policy's objective function $J(\theta)$ [48]. In other words, policy-based methods learn how these parameters should change the probabilities by which different actions can be taken in different states in order to maximize the expected reward. The main advantage of policy-based methods is their effectiveness for continuous action or the high dimensional space, where the parameters of the 'parameterized policy' are adjusted instead of solving a complicated maximization in every step. The policy gradients (PG) algorithm is widely used to solve the problems of the continuous action space. The policy is represented by a parametric probability distribution (see equation 6). In the PG algorithm, the action (a) at state (s) is selected stochastically based on a vector of parameters (θ), and by adjusting these parameters, the policy is driven in the direction of increasing the cumulative reward [49]. Policy gradient is the derivatives

(vector of derivatives) of the policy's objective function $J(\theta)$ with respect to the parameters (θ) as shown in equation 7 [51]. The problem can be formalized as shown in equation 8, considering (τ) is the agent's trajectory, $R(\tau)$ is the corresponding reward, (π_θ) is the parameterized policy and $P(\tau \mid \theta)$ is the probability of the trajectory (τ) under the policy (π_θ). The policy gradients algorithm searches for the local maximum by ascending the gradient of the policy with respect to the parameters (θ). It seeks to increase the probabilities of the trajectories that give the best return, as shown in equation 9. By reformulating the probability of the trajectory $P(\tau \mid \theta)$ and decomposing the trajectory into (states – actions), the policy gradients equation can be reformulated as shown in equation 10. Instead of integrating over the spaces of both state and action as in the case of stochastic policy gradients, deterministic policy gradients (DPG) integrates only over the state space, which in turns leads to a reduced number of samples, especially in the case of applications with large action states [48]. DPG is used in the deterministic environment (no uncertainty) where it accepts a state as input and outputs a single action $\pi_\theta(s) = a$. On the other hand, the stochastic policy is always needed to explore the complete state-action space. Based on that and for sufficient exploration for the DPG algorithm, the actions are chosen according to stochastic policy behavior, while learning a deterministic target policy. The policy that the agent uses to determine its actions at a given state is called behavior policy, while the policy that the agent uses to update the Q-value is called target policy. Learning the policy can be achieved in two different algorithms, on-policy or off-policy [52]. In the case of on-policy learning, the behavior policy is the same as the target policy, while they are different in the case of the off-policy learning algorithm.

$$\pi_\theta = P[a \mid s, \theta] \tag{6}$$

$$\nabla_\theta J(\theta) = \begin{bmatrix} \frac{\partial J(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial J(\theta)}{\partial \theta_n} \end{bmatrix} \tag{7}$$

$$\theta^* = \arg\max_\theta J(\theta) = \max_\theta \sum\nolimits_\tau P(\tau|\theta) R(\tau) \tag{8}$$

$$\nabla_\theta J(\theta) = E_\tau \left(\nabla_\theta \log P(\tau|\theta) R(\tau)\right) \tag{9}$$

$$\nabla_\theta J(\theta) = E_\tau \left(\nabla_\theta \log \pi_\theta P(s|t)\right) \tag{10}$$

2.2.3 Actor-Critic Algorithm

Actor–critic algorithms combine the benefits of both value-based and policy-based algorithms. The essential idea is that a value function approximator (critic) is used to explicitly estimate the action-value function instead of using the return. These algorithms deal with two different sets of parameters using two different approximators, the critic and the actor. The critic updates the action-value function

parameters, while the actor updates the policy parameters based on the direction that is suggested by the critic [53]. Actor-critic algorithms use an approximate policy gradient as described in equation 11, where the $Q_W(s,a)$ is the estimated cation-value function. Deep Deterministic Policy Gradient (DDPG) is a model-free, off-policy, actor-critic reinforcement learning algorithm that searches for the optimal policy that maximizes the cumulative long-term return for the continuous action environment. DDPG uses deep neural network-based approximators [44]. In the DDPG algorithm, the actor is used to approximate the optimal policy deterministically, which is unlike the stochastic policy, where the policy learns the probability distribution rather than actions. After the action is taken by the actor, the critic evaluates that action in order to determine whether the new state is better or worse than the expectation. That can be achieved by maintaining the Q-values of the taken actions towards the target Q-values. RL has been applied to a variety of autonomous driving tasks, [54], [55], [56], [57].

$$\nabla_\theta J(\theta) = E_{\pi_\theta}\left(\nabla_\theta \log \pi_\theta(s|a)\, Q_w(s,a)\right) \tag{11}$$

2.3 Supervised Learning Compared to Reinforcement Learning

Unlike RL methods where the agent learns by interacting with the environment without any supervision data, in supervised learning, the agent learns using labeled data sets. This means that the expert is explicitly guiding the model on how to act based on the labeled data. In deep neural networks, for example, and during training, the network approximates the future outputs for the observations and then compares them with the labeled ones in order to reduce the error. Supervised learning is mainly dedicated to dealing with two main categories of tasks, classification and regression, whereas RL deals with Markov's decision processes, policy learning, and value learning. The simplicity and the speed of the convergence during the training are the advantages of supervised learning compared to reinforcement, where convergence to the optimal policy can be slow so it requires intensive time. On the other hand, the efficiency of the supervised model is greatly affected by the comprehensiveness of the training data-set. In the case of nonlinear and complex systems such as driving system tasks, sufficient training data must be ensured in order to provide an efficient and generalizable model in all complex driving environments. The use of deep learning in a variety of fields has increased recently due to new powerful processing technologies that reduce the training time and improve performance. The deep neural network algorithm is a self-optimization algorithm and it has the ability to adopt a new scenario, which enables the developers to generalize the desired models. These features make deep learning suitable for control applications within dynamic and complex environments. The computational complexity and the advantages of the learning-based methods compared to classical MPC are presented in the additional material, Section 2.3. (See: [1])

3 Design of the Controllers

This section includes the designing process of the MPC, the DNN, the DQN and the combined models. The DNN model is developed to imitate the behavior of the MPC. The deep network of the DQN model will be designed with the same structure as that of the DNN model. In the combined model, the trained DNN model is combined with a Reinforcement DDPG algorithm as a decision maker. DQN and Hybrid models were trained until the determined criteria are achieved (desired reward, number of episodes, ... etc.)

3.1 Design of the MPC Controller

Since the MPC is a model-based controller, the first step in the design process is to design the vehicle model. Figure 2 shows the global position of the vehicle, while equations 12, 13 and 14 are the mathematical representation of the vehicle dynamic. Figure 3 shows the MPC model, while the input-output signals, the parameters, and the constraints of the MPC are presented in Table 1. During the designing process, the parameters of the MPC are initiated based on standard recommendations and were tuned during the testing until a stable behaviour is achieved. A detailed explanation about MPC optimization problem, Performance specifications, control law, and parameter calculations can be found in [1], section (3.1).

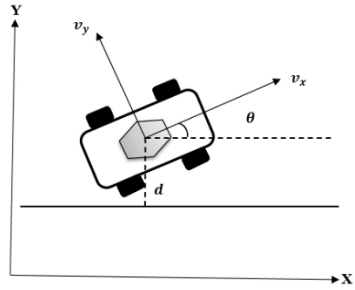


Figure 2
Global position of the vehicle

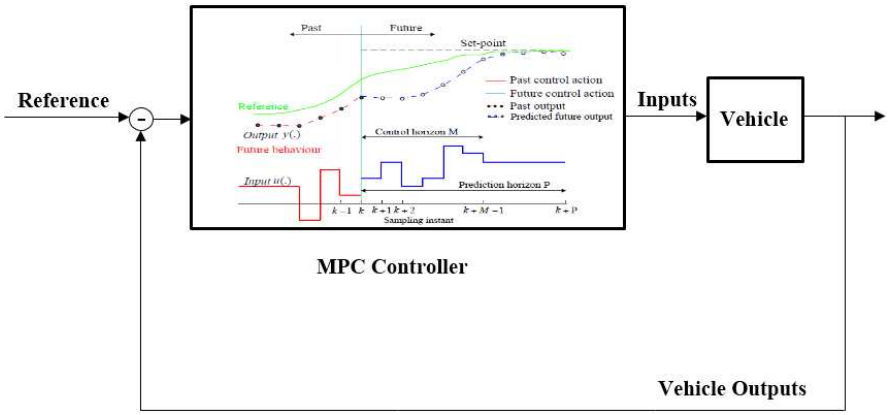


Figure 3
MPC controller design

$$\frac{d}{dt}\begin{bmatrix} v_y \\ \omega \\ d \\ \theta \end{bmatrix} = A\begin{bmatrix} v_y \\ \omega \\ d \\ \theta \end{bmatrix} + B\begin{bmatrix} \delta \\ \rho \end{bmatrix} \tag{12}$$

$$A = \begin{bmatrix} -\dfrac{2c_f + 2C_r}{mv_x} & -v_x - \dfrac{2c_f l_f - 2c_r l_r}{mv_x} & 0 & 0 \\ -\dfrac{2c_f l_f - 2c_r l_r}{l_z v_x} & -\dfrac{2C_f l_f^2 - 2C_r l_r^2}{l_z v_x} & 0 & 0 \\ 1 & 0 & 0 & v_x \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{13}$$

$$B = \begin{bmatrix} \dfrac{2c_f}{m} & 0 \\ \dfrac{2c_f l_f}{l_z} & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{14}$$

where v_x is longitudinal velocity; v_y is lateral velocity; d is lateral deviation; c_f is the corner stiffness of front tires; l_r is the distance between the rear tire and the center of the gravity; I_z is yaw moment; m is the vehicle's mass, ω is yaw rate; δ is steering angle; θ is yaw angle; c_r is the corner stiffness of rear tires; l_f is the distance between the front tire and the center of gravity; and ρ is the curvature.

Table 1
Design parameters and system constraints of the MPC controller

Internal model (vehicle)	Input signals	Steering angle (δ)
		Disturbance ($\rho\ v_x$)
	Output signals	Lateral deviation (d)
		Yaw angle (ω)
		Lateral velocity (v_y)
		Yaw rate (ω)
Parameters of MPC model	Sample time (T_s)	0.1 seconds
	Prediction horizon (P)	2 seconds
	Control horizon (M)	2 seconds
Constraints	Steering angle	[-1.04, 1.04] rad
	Changing rate	[-0.26, 0.26] rad

3.2 Design of the DNN model Using Imitation Learning

To achieve imitation learning, the DNN model was designed and structured based on the MPC model, where six observations are determined as inputs ($\theta,\ v_y, d, \omega, \rho,$ $\hat{\delta}$) and one control action (δ) was determined as an output, where ($\hat{\delta}$) is the previous control action. The detailed structure is shown in Figure 4. In regard to the training options, Adaptive Moment Estimation (ADMA) is used as an optimizer, the maximum number of Epoch is set to be 40, the mini-batch for each iteration is set to be 420, and the initial learning rate is set to be 0.01. Data preparation and training process of the MPC controller is presented in [1], Section 3.2.1

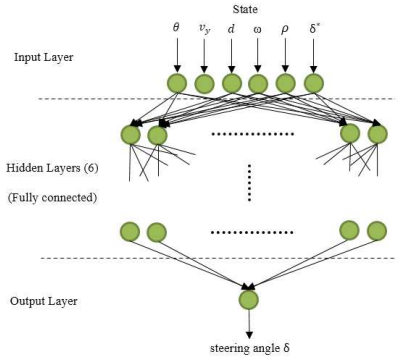


Figure 4
The DNN model structure

3.3 Design of the Reinforcement Deep Q-Learning Model

The desired DQN model is designed taking into consideration the same dynamics of the vehicle, the constraints, and the environment conditions that were used previously. The designing processes went through several steps, preparing the environment, creating and training the agent and finally testing and evaluating the performance. The environment is created using the six observations and the control action space was determined as a discrete space in the range of [-1.04, 1.04] rad, meaning that the agent can apply 121 possible actions at each state. Based on that, the deep Q-network is designed to accept the state from the environment as an input (vector with 6 observations) and outputs the estimated Q-values of each possible discrete action that can be taken at that state (vector of n=121 Q values). The detailed structure of the DQN model is shown in Figure 5. The target DQN, which is used to calculate the target Q-values is a clone of the DQN with the same structure and parameterization. The training details of the DQN model are presented in [1] Section 3.3.1.

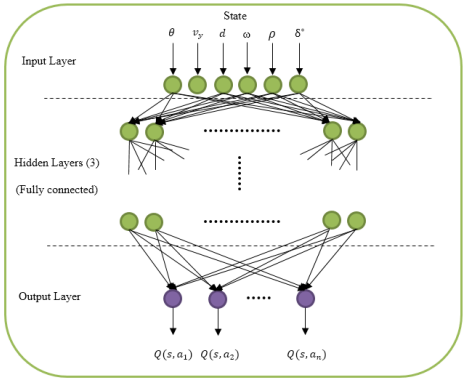


Figure 5
The reinforcement DQN structure

3.4 Design of the Combined (Supervised–RL) Model

The same vehicle's dynamics, constraints, environment conditions and state space are used to design and test the combined model. The continuous actions space is determined to be in the range of [-1.04, 1.04] rad. In order to create the agent, beside having the trained DNN model as an actor, the critic is created based on the actions-observations specifications, where the neural network is structured to accept two inputs (state-action) and one output (the corresponding expected long-term reward $Q\ (s, a \mid \theta^Q)$, and 3 hidden layers. Figure 6 shows the detailed structure of the combined model. The training process is presented in Section 3.4.1 of [1]

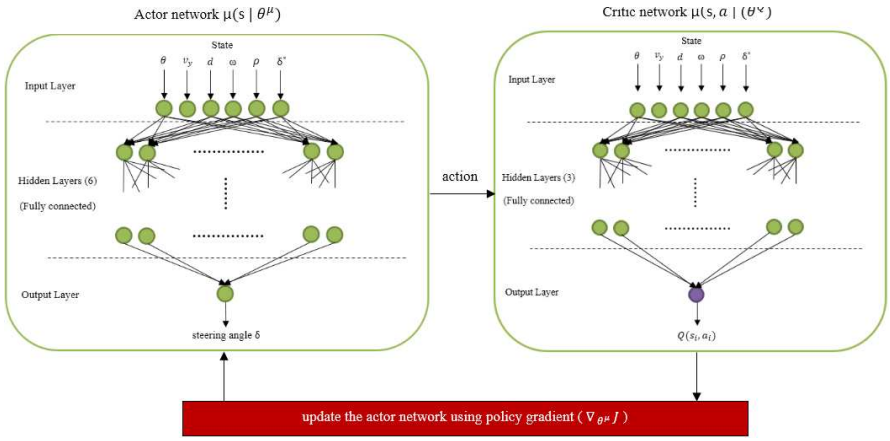


Figure 6
The structure of the actor-critic networks - combined model

4 Results and Discussion

The implementations of the designed models were performed using the same vehicle model and subjected to the same constraints, environmental conditions, and initial state. The performance is analyzed and evaluated taking into consideration the performance of the MPC controller as a reference behavior. The efficiency is discussed based on different indicators: the ability of the controllers to drive the vehicle along the desired trajectory in the first place, the time needed to reach a stable state, and the smoothness of the driving system. The obtained results in Figure 7 clearly show that the trained DNN and the MPC controller behave similarly with very small output deviation, where the maximum difference is approximately 0.0094 rad (0.53 degrees). The behavior is evaluated based on the response of the vehicle to the controllers' outputs. Figure 8 shows vehicle response to the control actions of the MPC and the DNN models in terms of lateral deviation and Figure 9 shows that both controllers (MPC-DNN) were able to follow the desired trajectory by driving the lateral deviation and yaw angle to be very close to zero. Additionally, and taking into consideration the control system characteristics, the results clearly show that both controllers were able to reach the stable state at almost the same time with the same amount of overshooting. These results prove that the trained DNN model was able to imitate the behavior of the traditional MPC controller successfully. Based on that, the performances of the reinforcement DQN model and combined model are compared to the DNN model in order to evaluate the best result achieved by the machine-learning-based models. Figure 10 shows that the three models responded differently to the same initial state. Despite these differences, Figures 11 and 12 show that the reinforcement DQN and the combined models were able to track the desired trajectory with different control system characteristics

(steady state time and overshooting). The detailed results showed that the combined model responded in a way that improved the smoothness of the driving system by reducing the overshooting (with hardly any overshooting in the case of lateral deviation) and drove the lateral deviation to be very close to zero (0.003 m) in a reasonable time, compared to the DNN model which achieved 0.0009 m as a final value of the lateral deviation at almost the same time but with higher overshooting and thus higher lateral deviations. The DQN model was not as efficient as the other models; its behavior led to higher overshooting and drove the lateral deviation to a final value of 0.01 m. As a result, and taking all the performance indicators into considerations, one can state that the combined model provided the best result and achieved the expected optimization by demonstrating accurate control actions (steering angles) that steer the vehicle along the desired trajectory efficiently in a reasonable time and improve the robustness of the driving system, while the DQN model, which is completely based on an RL algorithm, was not as efficient as the other two models (the supervised DNN or the combined model). The promising results that are provided by the reinforcement learning methods (DQN and Hybrid model) emphasize the importance of devoting more efforts to transferring them into practice as an efficient alternative to classical control methods.

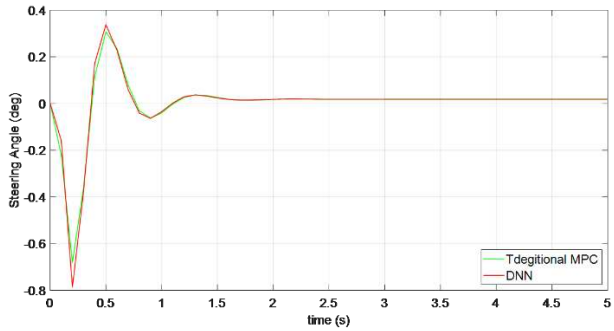


Figure 7
Comparison of the estimated steering angles of the MPC and the DNN models

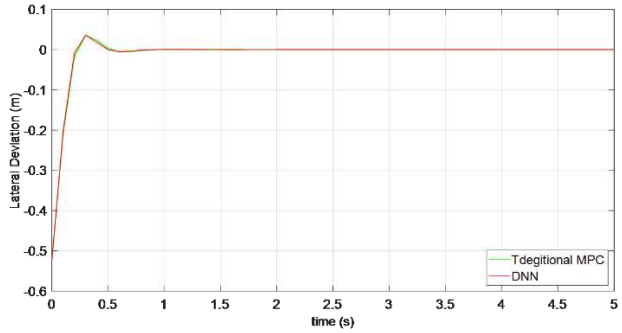


Figure 8
Vehicle response to the control actions of the MPC and the DNN models - lateral deviation

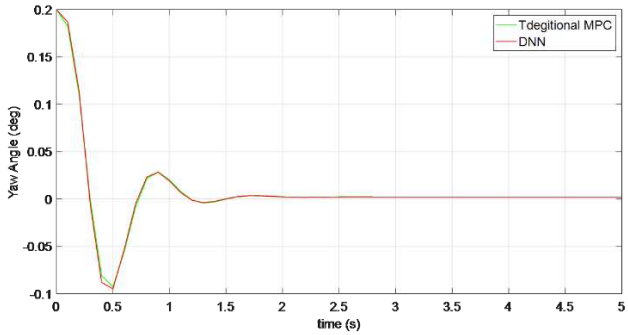


Figure 9
Vehicle response to the control actions of the MPC and the DNN models - yaw angle

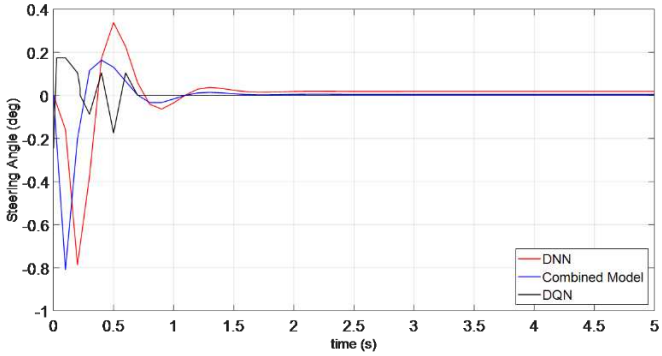


Figure 10
Comparison of the estimated steering angles of the DNN, DQN, and combined models

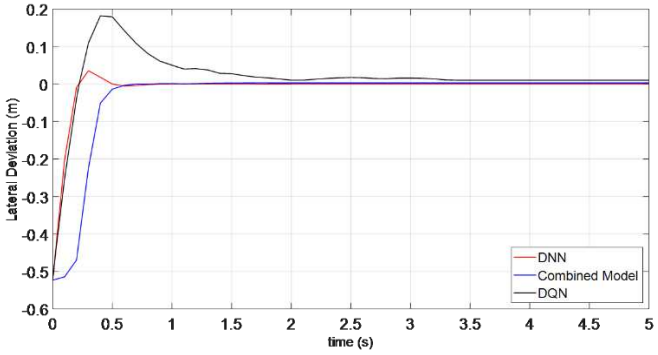


Figure 11
Vehicle response to the control actions of the DNN, DQN - lateral deviation

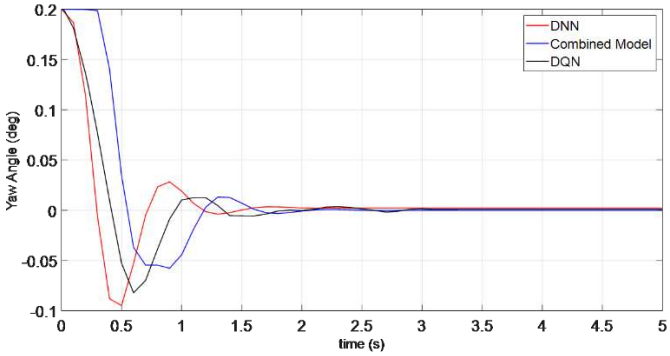


Figure 12
Vehicle response to the control actions of the DNN, DQN - yaw angle

Conclusions

In this work three different machine learning-based models were designed to perform an automated path-tracking task: a DNN model to imitate the behavior of the traditional MPC controller, a reinforcement learning DQN model, and a hybrid model. The hybrid model was designed to optimize the performance by combining the trained DNN model with the reinforcement learning model, where the DNN network was used as a decision-maker along with the critic network that evaluates the actions taken. The results showed that all three models were able to drive the vehicle along the desired path. The combined model was able to provide the desired optimization by driving the vehicle to the reference speed more smoothly and within a reasonable time. This work shows the efficiency of combining supervised and reinforcement learning to leverage the advantages of both algorithms, where the supervised learning speeds up the learning process and the reinforcement learning improves self-adaptation to new states that the model was not faced within the training process, which increases efficiency in the complex driving environment.

References

[1] A. Reda, R. Benotsmane, A. Bouzid, J. Vásárhelyi: Additional material for a Hybrid Machine Learning-Based Control Strategy for Autonomous Driving Optimization paper Acta1034; submitted to Acta Polytechnica Hungarica ISSN, https://drive.google.com/drive/folders/14hjWFoAxhHnWi18W-iO23QZe2btwF-de, See: Additional-material-ACTA-1034.pdf, 2023, p. 11

[2] C. Thorpe, M. Herbert, T. Kanade, S. Shafter: Toward autonomous driving: the CMU Navlab. II. Architecture and systems, in IEEE Expert, Vol. 6, No. 4, 1991, pp. 44-52

[3] E. Dickmanns, A. Zapp: Autonomous high speed road vehicle guidance by computer vision 1, IFAC Proc. Volumes, Vol. 20, No. 5, 1987, pp. 221-226

[4] SAE: On-Road Automated Vehicle Standards Committee, Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems, Technical Report, 2014, pp. 1-6

[5] U.S Department of Transportation: Preparing for the Future of Transportation: Automated Vehicles 3.0, 2018, https://www.transportation.gov/av/3, last visited December 2022

[6] B. Aelbachir, J. C. Smal, J. C. Blosseville, D. Gruyer: Simulation-driven validation of advanced driving-assistance systems, Procedia-Social and Behavioral Sciences, Vol. 48, 2012, pp. 1205-1214

[7] S. Hima, S. Glaser, A. Chaibet, B. Vanholme: Controller design for trajectory tracking of autonomous passenger vehicles, 4th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2011, pp. 1459-1464

[8] X. Li, Z. Wang, J. Zhu, Q. Chen: Adaptive tracking control for wheeled mobile robots with unknown skidding, 2015 IEEE Conference on Control Applications (CCA), 2015, pp. 1674-1679

[9] X. Wu, P. Lou, D. Tang, J. Yu: An intelligent-optimal predictive controller for path tracking of Vision-based Automated Guided Vehicle, 2008 International Conference on Information and Automation, 2008, pp. 844-849

[10] C. Sun, X. Zhang, Q. Zhou, Y. Tian: A Model Predictive Controller With Switched Tracking Error for Autonomous Vehicle Path Tracking, in IEEE Access, Vol. 7, 2019, pp. 53103-53114

[11] J. B. Rawlings, D. Q. Mayne: Model Predictive Control: Theory and Design, Nob Hill Publishing, Madison, 2009

[12] M. Brown, J. Funke, S. Erlien, J. C. Gerdes: Safe driving envelopes for path tracking in autonomous vehicles, Control Engineering Practice, vol. 61, 2017, pp. 307-316

[13] A. Reda, A. Bouzid, J. Vásárhelyi: Model predictive control for automated vehicle steering, Acta Polytechnica Hungarica, Vol. 17, 2020, pp. 163-182

[14] A. Krizhevsky, I. Sutskever, G. E. Hinton: ImageNet classification with deep convolutional neural networks, Commun. ACM, Vol. 60, No. 6, 2017, pp. 84-90

[15] L. Deng, J. Li, J. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams, Y. Gong, A. Acero, Recent advances in deep learning for speech research at Microsoft, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 8604-8608

[16] S. Kuutti, R. Bowden, Y. Jin, P. Barber, S. Fallah: A Survey of Deep Learning Applications to Autonomous Vehicle Control, in IEEE Transactions on Intelligent Transportation Systems, Vol. 22, No. 2, 2021, pp. 712-733

[17] M. Pfeiffer, M. Schaeuble, J. Nieto, R. Siegwart, C. Cadena: From perception to decision: A data-driven approach to end-to-end motion planning for autonomous ground robots, In 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 1527-1533

[18] R. Girshick, J. Donahue, T. Darrell, J. Malik: Rich feature hierarchies for accurate object detection and semantic segmentation, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580-587

[19] M. Batta: Machine learning algorithms - a review. International Journal of Science and Research (IJSR), 2020, pp. 381-386

[20] Y. Li, H. Li, Z. Li, H. Fang, A. Sanyal, Y. Wang, Q. Qiu: Fast and Accurate Trajectory Tracking for Unmanned Aerial Vehicles based on Deep Reinforcement Learning, 2019 IEEE 25th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), 2019, pp. 1-9

[21] S. Gu, E. Holly, T. Lillicrap, S. Levine: Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates, IEEE International Conference on Robotics and Automation, 2017, pp. 3389-3396

[22] A. B. Martinsen, A. M. Lekkas: Curved Path Following with Deep Reinforcement Learning: Results from Three Vessel Models, OCEANS 2018 MTS/IEEE Charleston, 2018, pp. 1-8

[23] S. Bacha, M. Y. Ayad, R. Saadi, A. Aboubou, M. Bahri, M. Becherif: Modeling and control technics for autonomous electric and hybrid vehicles path following, 2017 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B), 2017, pp. 1-12

[24] G. V. Raffo, G. K. Gomes, J. E. Normey-Rico, C. R. Kelber, L. B. Becker: A Predictive Controller for Autonomous Vehicle Path Tracking, IEEE Transactions on Intelligent Transportation Systems, Vol. 10, No. 1, 2009, pp. 92-102

[25] J. Wang, J. Steiber, B. Surampudi: Autonomous ground vehicle control system for high-speed and safe operation, International Journal of Vehicle Autonomous Systems, Vol. 7, No. 1, 2009, pp. 18-35

[26] Z. Sun, Q. Chen, Y. Nie, D. Liu, H. He: Ribbon Model based path tracking method for autonomous land vehicle, 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012, pp. 1220-1226

[27] A. De Luca, G. Oriolo, C. Samson: Feedback control of a nonholonomic car-like robot, In Robot motion planning and control, Springer, Berlin, 1998, pp. 171-253

[28] N. H. Amer, H. Zamzuri, K. Hudha, Z. A. Kadir: Modelling and control strategies in path tracking control for autonomous ground vehicles: A review

of state of the art and challenges, Journal of Intelligent $\&$ Robotic Systems, Vol. 86, No. 2, 2017, pp. 225-254

[29] E. Lucet, R. Lenain, C. Grand: Dynamic path tracking control of a vehicle on slippery terrain, Control Engineering Practice, Vol. 42, 2015, pp. 60-73

[30] C. Poussot-Vassal, O. Sename, L. Dugard, S. M. Savaresi: Vehicle dynamic stability improvements through gain-scheduled steering and braking control, Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility, Vol. 49, No. 10, 2011, pp. 597-1621

[31] P. Zhao, J. Chen, T. Mei, H. Liang: Dynamic motion planning for autonomous vehicle in unknown environments, IEEE Intelligent Vehicles Symposium (IV), 2011, pp. 284-289

[32] F. N. Martins, W. C. Celeste, R. Carelli, M. S.-Filho, T. F. B.-Filho: An adaptive dynamic controller for autonomous mobile robot trajectory tracking, Control Engineering Practice, Vol. 16, No. 11, 2008, pp. 1354-1363

[33] X. Ji, X. He, C. Lv, Y. Liu, J. Wu: Adaptive-neural-network-based robust lateral motion control for autonomous vehicle at driving limits, Control Engineering Practice, Vol. 76, 2018, pp. 41-53

[34] C. Sun, X. Zhang, L. Xi, Y. Tian: Design of a path-tracking steering controller for autonomous vehicles, Energies, Vol. 11, No. 6, 2018, p. 1451

[35] A. J. Babqi and B. Alamri: A comprehensive comparison between finite control set model predictive control and classical proportional-integral control for grid-tied power electronics devices, Acta Polytechnica Hungarica, Vol. 18, No. 7, 2021, pp. 67-87

[36] R. E. Precup, S. Preitl, E. M. Petriu, J. K. Tar, M. L. Tomescu, C. Pozna: Generic two-degree-of-freedom linear and fuzzy controllers for integral processes, J Franklin Inst, Vol. 346, No. 10, 2009, pp. 980-1003

[37] Zs. Preitl, R. E. Precup, J. K. Tar, M. Takács: Use of Multi-parametric Quadratic Programming in Fuzzy Control Systems, Acta Polytechnica Hungarica, Vol. 3, No. 3, ISSN 1785-8860, 2006, pp. 29-43

[38] T. Chen, A. Babanin, A. Muhammad, B. Chapron, and C. Chen: Modified Evolved Bat Algorithm of Fuzzy Optimal Control for Complex Nonlinear Systems, ROMJIST, Vol. 23, No. 672, 2020, pp. 28-40

[39] H. Redjimi, J. K. Tar: Multiple Components Fixed Point Iteration in the Adaptive Control of Single Variable 2nd Order Systems, Acta Polytechnica Hungarica, Vol. 18, No. 9, 2021, pp. 69-86

[40] A. Reda, J. Vásárhelyi: Model-Based Control Strategy for Autonomous Vehicle Path Tracking Task, Acta Universitatis Sapientiae, Electrical and Mechanical Engineering, Vol. 12, No. 1, 2020, pp. 35-45

[41] A. Zamfirache, R. E. Precup, R. C. Roman, and E. M. Petriu: Policy Iteration Reinforcement Learning-based control using a Grey Wolf Optimizer algorithm, Inf Sci, Vol. 585, 2022, pp. 162-175

[42] S. Sharifzadeh, I. Chiotellis, R. Triebel, D. Cremers: Learning to drive using inverse reinforcement learning and deep Q-networks, arXiv preprint arXiv:612.03653, 2016, Available: http://arxiv.org/abs/1612.03653

[43] M. L. Puterman: Markov decision processes: discrete stochastic dynamic programming, John Wiley $\&$ Sons, New York, 1994

[44] R. A. Howard: Dynamic Programming and Markov Processes, Cambridge, MA, USA: MIT Press, 1960

[45] R. S. Sutton, A. G. Barto: Reinforcement Learning: An Introduction, Cambridge, MA: MIT Press, 1998

[46] V. Francois-Lavet, P. Henderson, R. Islam, M. G. Bellemare, J. Pineau: An Introduction to Deep Reinforcement Learning, Foundations and Trends in Machine Learning, Vol. 11, No. 3-4, 2018, pp. 219-354

[47] Y. Li: Deep reinforcement learning: An overview, arXiv preprint arXiv:1701.07274, 2017

[48] K. Arulkumaran, M. P. Deisenroth, M. Brundage, A. A. Bharath, Deep Reinforcement Learning: A Brief Survey, in IEEE Signal Processing Magazine, Vol. 34, No. 6, 2017, pp. 26-38

[49] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra: Continuous control with deep reinforcement learning, ICRL, 2016

[50] V. Mnih et al.: Human-level control through deep reinforcement learning, Nature, Vol. 518, No. 7540, 2015, pp. 529-533

[51] R. S. Sutton, D. A. McAllester, S. P. Singh, Y. Mansour: Policy gradient methods for reinforcement learning with function approximation, Advances in Neural Information Processing Systems, 2000, pp. 1057-1063

[52] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, M. Riedmiller: Deterministic policy gradient algorithms, In International conference on machine learning, 2014, pp. 387-395

[53] J. C. Jesus, J. A. Bottega, M. A. S. L. Cuadros, D. F. T. Gamarra: Deep Deterministic Policy Gradient for Navigation of Mobile Robots in Simulated Environments, International Conference on Advanced Robotics (ICAR), 2019, pp. 362-367

[54] A. E. Sallab, M. Abdou, E. Perot, S. Yogamani: End-to-end deep reinforcement learning for lane keeping assist, MLITS, NIPS Workshop, Vol. 2, 2016

[55] P. Wang, C.-Y. Chan, A. de La Fortelle: A reinforcement learning based approach for automated lane change maneuvers, IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018, pp. 1379-1384

[56] Z. Huang, X. Xu, H. He, J. Tan, Z. Sun: Parameterized batch reinforcement learning for longitudinal control of autonomous land vehicles, IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 49, No. 4, 2017, pp. 730-741

[57] H. Chae, C. M. Kang, B. Kim, J. Kim, C. C. Chung, J. W. Choi: Autonomous braking system via deep reinforcement learning, IEEE 20th International conference on intelligent transportation systems (ITSC), 2017, pp. 1-6

Optimization-based Re-routing and Reserve Activation Dispatch Calculations in Natural Gas Network Models

Dávid Csercsik[1,2]

[1] Centre for Economic and Regional Studies, Institute of Economics, Tóth Kálmán u. 4., H-1097 Budapest, Tel.: +36-1 309 26 52, Fax: +36-1 319 31 36
Email: csercsik.david@krtk.hun-ren.hu
[2] Pázmány Péter Catholic University, Faculty of Information Technology and Bionics, H-1083 Budapest, Práter u. 50/A

Abstract: Computational models corresponding to supply security in natural gas networks aim to describe flows and consumption values in the case of component failures or unforseen pipeline shutdowns. The role of natural gas reservoirs in this process has only been marginally analyzed in such models, and typically only on the level of countries, not on the level of international networks. In this paper, a computational framework is proposed to determine the resulting flows of a natural gas network following a line failure, considering the potential re-routing of available sources and reservoir activations as well. The proposed dispatch-type model is capable of explicitly considering real, and publicly available flow and reservoir data to determine the resulting flows in the case of component outages, considering the re-routing process as well.

Keywords: Natural gas networks, supply security, flow models, optimization;

1 Introduction

Investments and developments related to natural gas infrastructure are considered as top priority projects of significant cost for every country. The two main aspects taken into account during the planning of natural gas network or reservoir developments are (I) the expected economic benefits of the project, and (II) the potential effects regarding the security of supply.

The expected economic effects [1], the short and long term returns of gas infrastructure developments can be estimated only in the context of the complex, regional financial and geopolitical environment (in other words, the international market for natural gas). At first glance it may be interesting, but economic benefits of a new pipeline are not necessarily related to the realized physical transport on the new transport route. The Velke Zlievce interconnector between Hungary and Slovakia for example [2] allows Hungary to access western-European trading plattforms

(HUBs). This potential access to alternative sources improved the bargaining positions of Hungary and resulted in the lower prices of Russian gas export to Hungary. The physical usage of the new interconnector is practically negligible ever since, but its economic effects are significant. Several non-cooperative [3, 4] and cooperative strategic models [5, 6] aim to describe the economic aspects of infrastructure developments related to natural gas networks.

In contrast to economic analysis, which (typically) assumes normal operation of the infrastructure, the perspective of supply security focuses on scenarios, when the operation of the infrastructure and the network flows are negatively affected by external factors [7]. The underlying causes may be of technical nature (as failures of pipeline elements or compressor stations), but they may be related to political disputes as well, as in the case of the 2006 and 2009 Ukrainian-Russian gas crises [8, 9].

Yergin [10] summarized the principles required to ensure a secure energy network: diversification of supply, reserve supply, integrity of network and importance of information.

The importance of the diversification of supply is evident. In the case of supply disruption, the negative effects on the network are reduced if there are multiple sources or/and if there are alternative routes to the available sources. Recent reports indicate that diversification efforts of the EU in the past decades seem insufficient, and, in addition to the already experienced price surges, the implications of the Russian-Ukrainian war will potentially affect the supply security of Europe negatively also on the level of available gas reserves [11].

The next principle states the relevance if reserve supply which helps to create a flexible system. In the network different elements can serve as reserve including spare production capacity, gas storages close to the consumption or along the supply chain. The third principle emphasizes the integrity of the network. Networks that consist of separate territorial units must act as a single logistical unit in order to secure the supply e.g. by coordinated re-routing of the available supply. By ensuring the supply security together, the participants will be in a better position than if they do not take the others into account. The last principle states that the information flow between the network elements must be efficient so that the network can act together against the disruption. The aim of international energy organisations is to provide transparency and information for the participating nations.

In contrast to economic analyses, some aspects of supply security may be studied on the level of physical flows, not considering the prices of the resources and infrastructure – such approaches are known as dispatch type models. In the case of supply security events, when the consumption of certain network nodes is reduced due to disruptions in the transportation, we will distinguish two important elements in the restoration process. The first is the potential re-routing of gas available from sources already used during the normal operation of the network, and the other is the activation of additional sources as gas reservoirs or previously unused LNG terminals in order to mitigate the damage.

Although the literature related to supply security of natural gas is quite extensive

– for publications relevant for the European region one may refer to [12–14] – , studies using computational modelling tools and numerical simulations are less widespread [15, 16]. Moreover, even among quantitative models related to supply security, there are only a few studies analyzing the role of gas reservoirs [17], and these papers typically study the problem on the level of countries [18], not on the level of the continental network. The importance of a gas reservoir in a supply-disruption event depends on multiple factors. First, in addition to the total capacity of the reservoir, the maximal outlet rate limits the quantity of gas which may be withdrawn from a reservoir during a given time period (e.g. in a week or month). Second, the gas withdrawn from the reservoir must be transported to the nodes, where the consumption needs to be restored, thus free pipeline capacity must be accessible for the transport. As the available capacity depends on pipeline capacities and baseline flows as well, the availability (and thus the importance) of a given gas reservoir depends on the actual operational state of the network as well.

Our aim in this paper is to develop quantitative methods in order to model the re-routing and reservoir-activation scenarios taking place during the restoration process.

As the current paper focuses on the physical flows in supply security scenarios, no prices and costs are considered in this study: Neither the price of the production/transport of natural gas, nor the price of disruption of consumption is taken into account. Only physical flows and their limiting factors (pipeline, source and reservoir capacities) are considered.

2 Methodology

In this section we describe the concept of the proposed network-flow oriented framework. We consider the effects of transport disruption, and restoration processes on the time scale of one month. This time interval may be regarded as relevant for computational supply-security studies, since after one month, technical problems causing disruptions in the network operations are usually relieved. In addition, this time detail also fits the data available on network flows, which are important inputs of the model in applications. The proposed computational principles may be however easily extended to longer or restrained to shorter periods.

2.1 Basic Concepts of the Model

The natural gas pipeline network is represented in the model as a directed acyclic graph with n nodes and m edges. If we assume that no flow direction may be altered and no counter directed flows are allowed to take place, the acyclic property of the network graph may be regarded as plausible, considering the current modelling aim and context.

The nodes of the network are characterized by the following quantities:

- Maximal monthly inlet value in million cubic meters (mcm): This value represents non-reservoir type sources of natural gas, as production sites (wells),

LNG terminals etc., assuming normal opration. This value is denoted by $\bar{I}_j$ for node j.

- Monthly potential reservoir inlet value (in mcm): This value represents reservoirs, from which natural gas may be withdrawn. This value (denoted by $\bar{R}_j$) is equal to the volume of natural gas, which may be withdrawn from the reservoir in node j.
- Nominal monthly consumption value (in mcm): This value represents the monthly gas consumption of the node under normal circumstances (no supply disruption). We assume that this variable also means an upper bound for consumption. This value is denoted by $\bar{C}_j$ for node j.

Each edge i of the network is characterized by a capacity value denoted by $\bar{f}_i$.

2.1.1 Example Network

As an example (motivated by the simple network described in [15]), let us consider a 6-node ($n = 6$) network, depicted in Fig 1.

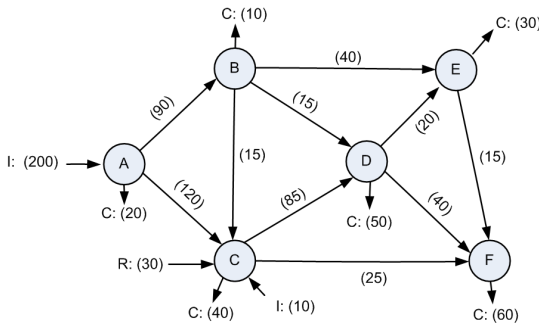


Figure 1
Simple example network with 6 nodes. *I* corresponds to inlets: the numbers in parentheses correspond to the maximal monthly inlet value of the respective node ($\bar{I}_j$). *R* corresponds to reservoirs: the numbers in parentheses correspond to the monthly potential reservoir inlet value of the respective node ($\bar{R}_j$). *C* corresponds to consumption: the numbers in parentheses correspond to the nominal monthly consumption value of the respective node. Numbers in parentheses on edges correspond to maximal capacity ($\bar{f}$). Only nonzero values are included in the figure.

We denote the nodes of the network by A, B, C, D, E and F respectively. We also summarize the nodal parameters of the network in Table 1.

The network has 10 edges ($m = 10$), the parameters of which (capacity values) are summarized in Table 2.

2.1.2 Operation Modes of the Network

In the context of our modelling computations, we consider four different operation modes of a given network:

- *Normal operation mode* (NOM) describes the base case flows, inlets and consumptions in the network. We will assume that these values are defined prior.

Table 1
Node parameters of the example network.

index	node notation	$\bar{I}_j$	$\bar{R}_j$	$\bar{C}_j$
1	A	200	0	20
2	B	0	0	10
3	C	10	30	40
4	D	0	0	50
5	E	0	0	30
6	F	0	0	60

Table 2
Edge parameters of the example network .

index	from node	to node	$\bar{f}$
1	A	B	90
2	A	C	120
3	B	C	15
4	B	D	15
5	B	E	40
6	C	D	85
7	C	F	25
8	D	E	20
9	D	F	40
10	E	F	15

Let us note that there are publicly available databases, which may serve as reference, when one aims to apply the proposed model for a realistic network scenario. Regarding eg. the European natural gas network, one may refer on the one hand to the International Energy Agency [19] for production/consumption values and on the other hand to the AGSI+ Aggregated Gas Storage Inventory [20], regarding reservoir capacities and actual availability values. In addition, the data of the natural gas and liquefied natural gas flows in the European gas network is summarized in a table provided Gas Trade Flows data service [21], which collects data of the gas flows of the 31 participating countries. The aim of the GTF data service is to improve the transparency in natural gas networks. The provided data includes the exit and entry points of the pipeline, the trading countries, the maximal flow capacity and the monthly flow amount back to October 2008. The AGSI data set includes the European gas storages and the data is available on a daily basis. The daily data contains the amount of currently stored gas, the trend of the stored gas, the amount of injected and withdrawn gas and the maximal capacity of injection and withdrawal. The storage data can be viewed aggregated by country and the historical data is also available on the aggregated data. In other words, the proposed model is constructed in a data-driven approach, and it is able to directly interpret data corresponding to any period of interest.

- *Disrupted operation mode* (DOM) of the network will describe, how a disruption of a line or a node will affect the inlets, consumption values and flows

of the network.

- *Re-routed operation mode* (RROM) of the network aims to describe the first reaction to the disruption: Gas available from various inlets is re-routed on available paths to replace the gas volumes in nodes where the disruption caused decrease in the consumption.
- *Reserve-activated operation mode* (RAOM) describes the activation of reservoir sources, after the re-routing has taken place: Gas stored in reservoirs is unloaded and routed on the available line capacities to further mitigate the effects of the disruption.

State variables of the model: The state of the network in any given mode of operation is described with the following variables.

- Node-related variables: For each node j the actual inlet value ($I_j \leq \bar{I}_j$), actual reservoir inlet value ($R_j \leq \bar{R}_j$) and the actual consumption value ($C_j \leq \bar{C}_j$) are considered as state-variables. We only consider supply-security usage of reservoirs, thus we assume that reservoir inlets may be nonzero only in RAOM.
- Edge-related variables: Gas flows are represented by a flow vector $f \in \mathscr{R}^m$. We assume that every flow is nonnegative, which means that the direction of flows must coincide with the direction of the edge. We assume that the unit of vector f is million cubic meters (mcm). Furthermore $f \leq \bar{f}$.

For every operation mode, we require that inflow and outflow must be in balance for each node. After a motivational example, which demonstrates these network states and highlights the possible measures applicable for the significance of reservoirs in this model context, we will describe in detail how the DOM and the following RROM and RAOM may be calculated consecutively, based on the network parameters and on the prior given state variables corresponding to the NOM.

The aim of the current article is to propose an algorithm, according to which the state variables in the DOM, RROM and RAOM modes may be calculated, considering the network parameters, the (prior given) state variables of the NOM and the parameters of the considered line failure.

2.2 A Demonstrative Example of the Disruption-restoration Process

In this subsection we demonstrate how the 4 network states are realized, using the simple network depicted in Fig 1. Figure 2 depicts the 4 states of the network.

Fig. 2 (a) depicts the NOM of the network. As already mentioned, in the context of the paper we assume that the state variables corresponding to this state (nominal inlet, consumption and flow values) are given prior. We can see that in NOM, all consumption needs are fulfilled in the network, and nodal balances hold for each node.

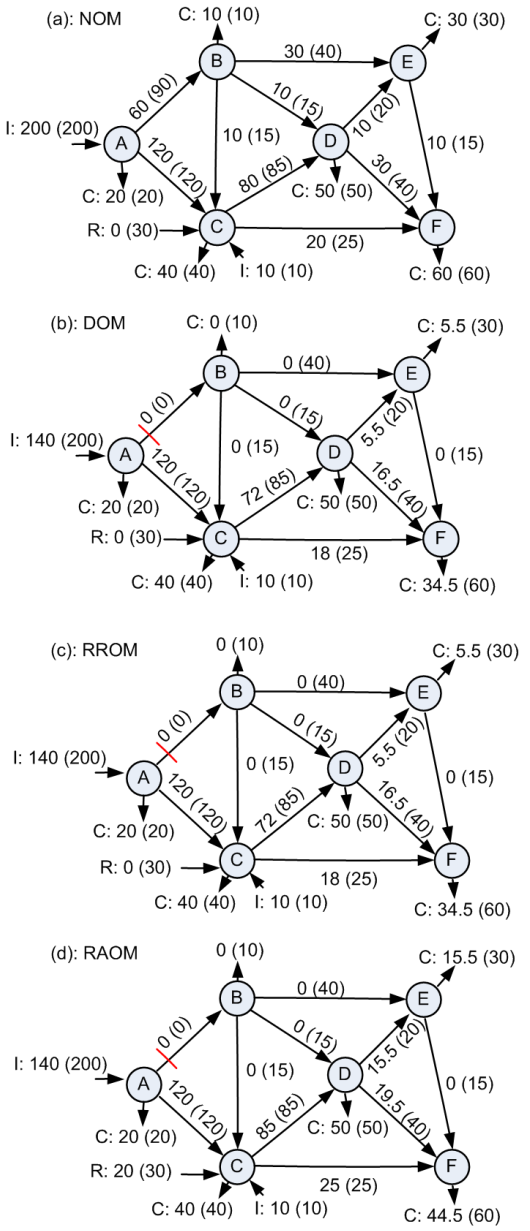


Figure 2
4 states of the example network. For each inlet (I), reservoir inlet (R) and consumption value (C), the actually realized values are indicated without parentheses, while the numbers in parentheses correspond to the maximal/ideal values (as before). The notation is the same for line flows: numbers without parentheses denote the realized values while numbers in parentheses are the maximal values (i.e. capacities).

Fig. 2 (b) depicts the DOM of the network. In this case, we assume the disruption of line 1 (between node A and node B). The inlet of A is reduced according to the amount originally transported on line 1 (60 units). Since no gas arrives in node B, it is straightforward that its outflows and consumption are zeroed out. In the case of nodes, which are also affected by the disruption, but still have incoming gas flows (like C, for which the input is decreased by 10 units), following [15], we assume that

- Each node prioritizes its own consumption: From the inflows the nodal consumption is covered first, and the rest is forwarded.
- The ratio of the forwarded gas volumes should be equal to the respective ratios in NOM.

Regarding node C, these assumptions imply that in the DOM, node C first covers its own consumption needs (40 units), and forwards the remaining gas to nodes D and F, according to the constraint regarding the proportions: $80/20 = 72/18$. The same applies for node D: First, the own consumption is covered and the remaining gas volumes are forwarded to nodes E and F according to the original proportions $(10/30 = 5.5/16.5)$.

Fig. 2 (c) depicts the RROM of the network. In this state, available (non-reservoir type) inlets are activated and the corresponding volumes are routed on free line capacities to compensate for the outages in the nodes affected by the disruption (in this case E and F). Let us note that during the model calculations we will assume that flows in the DOM must be unaffected by the re-routing – in other words, only available inlet volumes may be rerouted on available capacities (which are determined by the flows remaining in the DOM).

In this particular case however, only node A has available inlets, but the outgoing capacities of A are fully exploited, so this potential additional volume is not able to reach the nodes affected by the disruption (B, E and F). In other words, in the case of this example, no practical re-routing takes place,

Fig. 2 (d) depicts the RAOM of the network. In this case, the reservoir in node C is activated, and additional 20 units of gas is routed to E and F to compensate for the outage. The outward capacities of node C limit the activation of the reservoir (more gas can not be routed to E and F). We assume that the restoration volumes from the reservoirs are routed to achieve the most egalitarian compensation possible. In this case, perfect balance is possible (in the terms of node E and F, since B is not reachable from C): node E and node F both receive 10 units of gas from the reservoir in node C.

The total consumption outage implied by the considered disruption (after re-routing – which in his particular case implied no explicit flown modifications) was 60 units (10 units in B, 24.5 units in E and 25.5 units in F), from which the activation of the reservoir in node C was able to compensate 20 units. In this case, the reservoir of C compensated for 33.3 % of the consumption reduction.

2.2.1 Effect of Network Expansion

In this subsection we use the previously introduced simple example to show how network expansion can modify the modelled process, and the resulting measures. Let us assume that the nodal parameters of the network remain unchanged, but the capacity of the edges 2 (from A to C) and 6 (from C to D) are increased from 120 to 150 and 85 to 100 respectively. The updated edge parameters are summarized in Table 3, and the modified network is depicted in Fig 3.

Table 3
Edge parameters of the example network with extended capacities. Modified values are emphasized with bold typeface.

index	from node	to node	$\bar{f}$
1	A	B	90
2	A	C	**150**
3	B	C	15
4	B	D	15
5	B	E	40
6	C	D	**100**
7	C	F	25
8	D	E	20
9	D	F	40
10	E	F	15

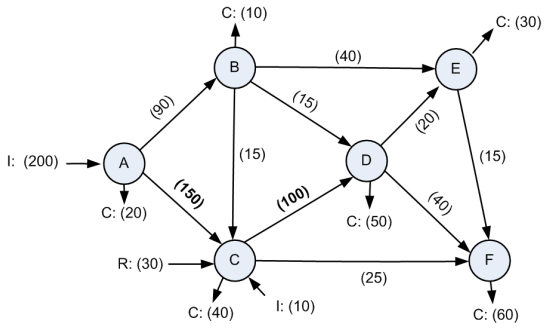


Figure 3
The simple example network with extended capacities. Modified values are emphasized with bold typeface.

We assume furthermore, that the NOM of the network is the same as before (depicted in Fig. 2 (a)). We consider again the same disruption, regarding line 1 (from A to B). The process, and the resulting 4 network states are depicted in Fig. 4.

In Fig. 4 (a) and (b) we can see that the base-case flows and the effect of the disruption are the same as before in Fig 2.

In Fig. 4 (c) however, in contrast to the original case depicted in Fig 2, we can see that re-routing has real significance in his case: In the RROM, an additional 30 units of gas from node A is re-routed to nodes E and F, using the available capacities of the

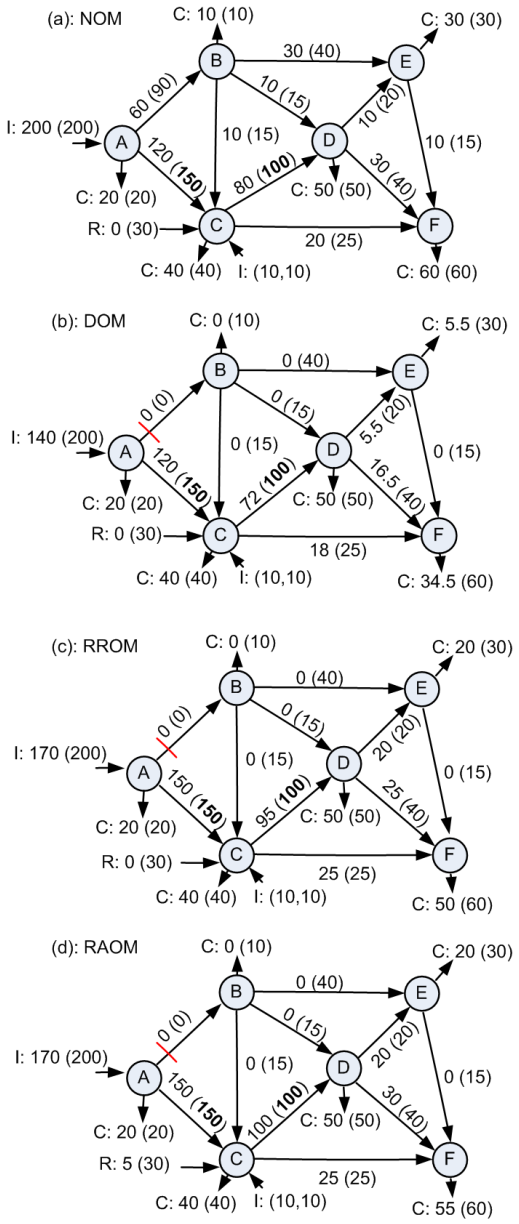


Figure 4

4 states of the extended example network. The parameters, which have been increased due to the extension are denoted with bold typeface. For each inlet (I), reservoir inlet (R) and consumption value (C), the actually realized values are indicated without parentheses, while the numbers in parentheses correspond to the maximal/ideal values (as before). The notation is the same for line flows: numbers without parentheses denote the realized values while numbers in parentheses are the maximal values (i.e. capacities).

network, which are present due to the capacity extensions. Although in subsection 2.3 we will see the detailed computational formulation of the re-routing process, the principles of the re-routing calculations are as follows.

We consider the nodes, where the DOM resulted in consumption outage, and which are reachable from the nodes with available additional inlet. We consider the outage values, which we aim to compensate, and look for a re-routing of available gas, which results in maximal overall compensation value (the sum of the remaining outages over the network must be minimal). As this solution, however, may not be unique, in the second step, from the possible outage-minimizing solutions regarding accessible outage nodes, we choose the one, which is the closest to 'equal compensation' in the absolute sense. As in general, this consideration does also not necessary imply a unique solution, in the third step we determine the setup which uses the shortest available routes.

In this particular case, 30 units are re-routed from node A. Node B is not accessible, thus we are interested only in nodes E and F. Node E has an outage of 24.5 units, while node F has an outage of 25.5 units. The equal compensation in the absolute sense would be if 15 - 15 units of the 30 units would be transferred to E and F. In this case, this is however not possible, so we choose the most close feasible solution: We transfer 14.5 to E via D and 15.5 to F directly.

Fig. 4 (d) depicts the activation of reserves. In this case, the free capacities allow 5 units of gas to be transferred from the reservoir of node C to node F.

In this case, we can say that the total consumption outage implied by the considered disruption (after re-routing – which in his particular case implied no explicit flown modifications) was 30 units (10 units each in B, E and F), from which the activation of the reservoir in node C was able to compensate 5 units. The reservoir of C compensated for 16.667 % of the consumption reduction in this case.

As we can see, the modification of the network affected (more precisely reduced) the significance of the reservoir in the context of its consumption-reduction potential. In the following we describe the details of the calculations corresponding to the determination of the network states corresponding to the various operation modes.

2.3 Details of the Computational Formulation

As we mentioned before, we assume that the NOM of the network is given prior. In the following we describe how the network states in further operation modes (DOM, RROM and RAOM) are derived from this initial state.

2.3.1 DOM Calculations

In the framework of the proposed model we consider line failures. Any line of the network may fail, and, according to our consumptions, this means that the maximal transport capacity of the line in question is reduced. Basically, we assume that the capacity of the line is reduced to 0, but partial failures may be also considered – the proposed computational methods may be applied in this case as well. This reduction

in the transfer capacity typically also induces the reduction of line flows, as in the case of Fig. 2 (b), where the flow of the line A-B has been reduced to 0.

This initial reduction of a given flow induces imbalances in some of the network nodes (i.e. the start node and end node of the edge in question). In order to determine the modified state variables (inlets, consumptions and line flows) of the network which resolve this imbalance and constitute the DOM state, we perform the following steps for each imbalanced node.

- If the node has surplus (the sum of inlets and inflows exceeds the sum of consumption and outflows), inlet and inflow values are decreased in a way which conserves the proportions of inlet and inflow values as much as possible.
- If the node has deficiency (sum of consumption and outflows exceeds the sum of inlets and inflows), consumption and outflow values are decreased. In this case however we assume that each node with deficiency gives priority to own consumption: Such nodes aim to cover own consumption first and the outflows are updated according to the remaining gas quantity.

After performing these steps, two possibilities may arise.

- There are no more imbalanced nodes in the network. In this case the calculations are finished, and the resulting states constitute the DOM.
- One or more imbalanced nodes are still present in the network. In this case, the calculations are repeated.

The above iterative process is demonstrated on a less trivial example in detail (step-by-step) in Fig. 5, where nodes with imbalance are highlighted. In this figure we can see that in the first step (b) the initial line failure affects the starting node and the end node of the failed line (namely E and G). In the next step (c), the balance of these nodes is restored according to the principles described above, but this implies further imbalance in nodes C and D. As the process progresses, in the end (f), all nodes are in balance again, and the network sates corresponding to the DOM are determined. The acyclicity of the graph guarantees that the process will stop after a finite number of iterations.

2.3.2 RROM and RAOM Calculations

The calculation of the states corresponding to RROM and RAOM is performed similarly, the only difference is that while in the case of RROM the unused capacity of (normal) inlets are considered as sources, in the case of RAOM the reservoir capacities are considered as sources. In the following we describe the RROM calculations – the RAOM can be calculated accordingly, mutatis mutandis.

If we compare the actual consumption values of the DOM with nominal consumption values we can determine the set of nodes (nodes with consumption outage), which are affected by the disruption (in Fig. 2 these are nodes C E and F). Following this we determine that subset of these outage nodes, which are reachable from any of the source nodes: An outage node (B) is reachable from the source node (A), if a directed path from A to B exists, along which all the lines have nonzero free

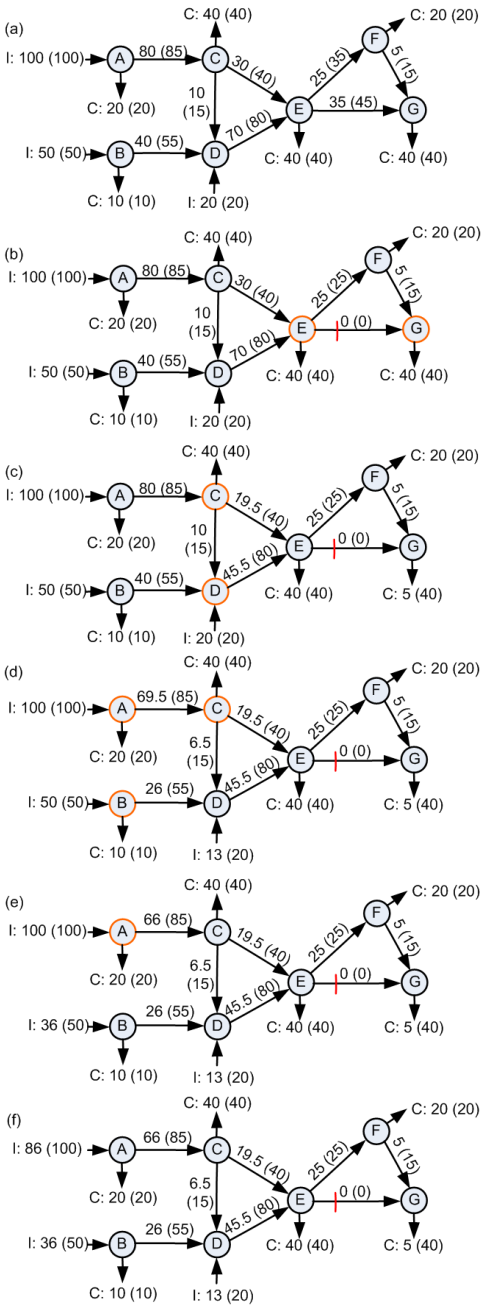


Figure 5
Iterative calculation of the DOM state in an example network. Imbalanced nodes are denoted with orange.

capacity (capacity over the actual flow). We denote the set of reachable nodes by N^R.

The calculation of the states corresponding to the RROM can be described as a sequence of flow-optimization processes. The variables of these flow-optimization problems are the same in each step (inlets of the sources, line flows and consumption values), but the constraints and the objective function is different in each step. The steps considered are as follows.

Maximizing the total restoration of consumption: In this step the following optimization problem is considered. For each affected node j (nodes for which the consumption has been decreased as a result of the line fault), the reduction in consumption can be defined as $C_j^- = C_j^{NOM} - C_j^{DOM}$ where C_j^{NOM} is the consumption of node j in NOM and C_j^{DOM} is the consumption of node j in DOM (the first is given prior, and the second is determined via the method described in subsection 2.3.1). The available additional inlets considered in the re-routing process are defined as $\bar{I}_j^{RR} = \bar{I}_j - I_j^{DOM}$ for each node j, where I_j^{DOM} denotes the inlet of node j in the DOM. The available line capacities for each line i in the re-routing process are defined as $\bar{f}_i^{RR} = \bar{f}_i - f_i^{DOM}$, where f_i^{DOM} denotes the flow of line i in the DOM.

The variable vector of the optimization corresponding to the first step of re-routing is defined as in Eq. (1).

$$x^{RR} = \begin{pmatrix} I^{RR} \\ C^{RR} \\ f^{RR} \end{pmatrix} \tag{1}$$

$I^{RR} \in \mathscr{R}^n$ stands for the additional inlets realized (activated) in the re-routing process, $C^{RR} \in \mathscr{R}^n$ denotes the vector of additional consumption values, resulting from the re-routing. These additional consumptions aim to mitigate the consumption-reductions resulting from the disruption. $f^{RR} \in \mathscr{R}^m$ denotes the flows corresponding to the re-routing process.

The constraints of the problem are described in equation (2), while eq. (3) formulates the objective function: The aim in this step is to restore as much consumption as possible. We can see that the optimization of this step results in a linear programming problem.

$$\begin{aligned} I_j^{RR} &\leq \bar{I}_j^{RR} \quad \forall\, j \\ C_j^{RR} &\leq C_j^- \quad \forall\, j \\ f_i^{RR} &\leq \bar{f}_i^{RR} \quad \forall\, i \end{aligned} \tag{2}$$

$$\max_{x^{RR}} \quad \sum_j C_j^{RR} \tag{3}$$

Let us denote the obtained maximal value of the objective function by C_{T1}^{RR}.

Equalizing the consumption-restoration over reachable nodes: As the set of x^{RR} vectors maximizing the total restoration of consumption may be unique, in the next step we formalize the consideration, that we prefer such vectors, which equally reduce the consumption reduction in each reachable node. This consideration is formulated as a quadratic programming problem, in which the variable vector is the same as before (see Eq. (1)), the constraints (2) still hold, but we add the constraint (4) as well, and modify the objective function as described in Eq. (5) where the indices i and j correspond to nodes in N^R.

$$\sum_j C_j^{RR} = C_{T1}^{RR} \tag{4}$$

$$\min_{x^{RR}} \quad \sum_{i \neq j} (C_i^{RR} - C_j^{RR})^2 \tag{5}$$

Let us denote the C^{RR} vector resulting from the optimization problem by C_1^{RR}.

Minimal usage of network lines: It is possible that the solution obtained for x^{RR} in the previous step is still not unique. In this case we add the constraint (6) to the problem, and minimize the linear objective function (7) to choose the flows which use the minimal number of edges in the network.

$$C^{RR} = C_{T1}^{RR} \tag{6}$$

$$\min_{x^{RR}} \quad f^R R \cdot 1 \tag{7}$$

The above consideration does not distinguish between the transportation costs among edges, but if transfer costs for the edges are available, this step may be modified accordingly.

3 Discussion

As it has been described in subsection 2.1, the proposed method assumes an acyclic flow pattern as input (corresponding to the NOM). This assumption of acyclicity deserves some discussion. It seems straightforward to assume that because of the transfer costs, cyclic flows are contra-productive and unnecessary in the network, so one may think that this assumption naturally holds for real data. In contrast, considering the realistic European pipeline network, the flows observable in any

instance are results of complex financial transactions composed of agreements for various terms, which are sometimes binding (e.g. it is possible that according to a long-term agreement gas must be transferred from east to west on a pipeline, even if on the actual gas is cheaper on western hubs). This may result in cyclic flows on the network. On the other hand, data available about pipeline gas flows is cumulated (e.g. for months). It is possible that daily acyclic flows result in monthly cumulated flows, which do have cycles. If the proposed computational framework is used to interpret real flow data, this issue may be solved by a preprocessing step of, which removes the cyclic flows from the input.

Regarding the practical applicability of the proposed method, the most computationally demanding step is the quadratic optimization described by eq. 5. While freely available solvers accessible e.g. by MATLAB may have a variable limit of hundreds, which may limit the size of the considered network model, commercially available solvers like CPLEX [22] are able to handle even millions of variables, if the memory of the applied hardware is sufficient.

As it has been already discussed, in the case described in subsection 2.2, the reservoir of C compensated for 33.3 % of the consumption reduction in the case of the failure of the first line, while in the case of the upgraded network, described in subsection 2.2.1, the reservoir was able to compensate 16.667 % of the outage, assuming the same line failure. If one calculates these values for all the possible line failures, the resulting values may serve as the basis of a significance measure of the reservoir in question. The most simple approach is to consider the average of these values as the supply security significance measure of the reservoir in question. Naturally, this significance value may not be calculated only for existing reservoirs, but also for potential new gas storage facility projets, in which case the proposed approach could represent a decision-support tool helping to evaluate the future effect of potential projects on supply.

In the current work we only addressed scenarios, in which a single reservoir was present in the network. In realistic cases, the storage capacities are distributed in the network, and in the case of a more serious outage event (line failure), they may be activated simultaneously to mitigate the effects on consumption reduction. Assuming such a scenario, the aforementioned concept of supply security significance measure has to be reconsidered under the assumption of potential co-activation of reservoir supplies.

Furthermore we have to emphasize that the model assumes a fully centralized process of re-routing and reserve distribution, which may be regarded only as a partially realistic assumption. While during normal operation of the international natural gas network, reservoir capacities, apart from the strategic reserves of individual countries, are dominantly marketed and allocated for energy trading companies, who use them for their own business-motivated purposes. In the case of serious outages however, like the 2009 Russo-Ukrainian dispute [9] these companies usually agree to handover their reserves for strategic damage control, allowing for more centralized decision making on the network level. Nevertheless, the filling of these reservoirs is usually not centrally coordinated, thus a EU-level strategic reserve including multiple reservoirs may significantly increase the preparedness for outage

events.

Conclusions and Future Work

In this paper, we provided a computational framework, which assigns supply security related significance measures to gas reservoirs. The method uses the concept of nominal, disrupted, re-routed and reserve-activated operation modes (NOM, DOM, RROM and RAOM respectively), which are calculated iteratively. The assumptions used during the calculations try to avoid the prioritization of any consumer or line, and they are formulated according to the principle that both the effect of failures and restoration efforts take place without any prioritization among consumers or network lines.

During the DOM calculations described in subsection 2.3.1, the effect of flow disruptions are iteratively back-propagated through the network, assuming proportional decrease of nodal inflows (we require the inflow proportions to be constant in every affected node). During the forth-propagation of flow disruptions, it is assumed that every node prioritizes its own consumption and sends the remaining gas further, aiming to keep the proportion of nodal outflows – this assumption is in accordance with [15].

During the RROM and RAOM calculations, the first principle is always to restore as much consumption as possible, and the second aim is to do this via the most equal support of affected nodes, which are reachable from additional sources.

According to the above calculations in the case of a line failure, we can determine how much of the consumption reduction may be restored based on the emergency activation of the reservoir in question. Averaging these values for every possible line failure considered results in a quntitative supply security related significance measure for the corresponding gas reservoir.

3.1 Future work

As discussed earlier, based on the proposed calculation algorithms, it is possible to define and calculate supply security measures for reservoirs of natural gas, in order to characterize their importance in the restoration process.

In addition, the acyclicity assumption regarding the NOM flows is a critical element of modelling assumptions. In the future, in addition to the iterative method described in subsection 2.3.1, it would be desirable to develop additional approaches, which may be used for the determination of DOM in a network, even in the case of cyclic flows present. An optimization-based approach could possibly substitute the iterative algorithm. The details of this alternative DOM calculation will be described in the future.

Acknowledgement

This work has been supported by the Hungarian Academy of Sciences under its Momentum Programme LP2021-2 and by the Fund K 131 545 of the Hungarian National Research, Development and Innovation Office.

References

[1] Jacek Brożyna, Grzegorz Mentel, and Beata Szetela. Renevable energy and economic development in the european union. *Acta Polytechnica Hungarica*, 14(7):11–34, 2017.

[2] Matúš Mišík and Andrej Nosko. The eastring gas pipeline in the context of the central and eastern european gas supply challenge. *Nature Energy*, 2(11):844–848, 2017.

[3] Franziska Holz, Christian Von Hirschhausen, and Claudia Kemfert. A strategic model of european gas supply (gasmod). *Energy Economics*, 30(3):766–788, 2008.

[4] Rudolf Egging-Bratseth, Franziska Holz, Alexandra Lüth, Sebastian Wegel, and Jan Zepter. Global gas model: Model and data documentation v3. 0 (2019). Technical report, DIW Berlin, German Institute for Economic Research, 2019.

[5] Franz Hubert and Svetlana Ikonnikova. Investment options and bargaining power: the eurasian supply chain for natural gas. *The Journal of Industrial Economics*, 59(1):85–116, 2011.

[6] Balázs R Sziklai, László Á Kóczy, and Dávid Csercsik. The impact of nord stream 2 on the european gas market bargaining positions. *Energy Policy*, 144:111692, 2020.

[7] Bülent Oral and Ferdun Dönmez. The impacts of natural disasters on power systems: anatomy of the marmara earthquake blackout. *Acta Polytechnica Hungarica*, 7(2):107–118, 2010.

[8] Jonathan Stern. The russian-ukrainian gas crisis of january 2006. *Oxford Institute for Energy Studies*, 16(1), 2006.

[9] Jonathan Stern, Simon Pirani, and Katja Yafimava. *The Russo-Ukrainian gas dispute of January 2009: a comprehensive assessment*. Oxford Institute for Energy Studies, 2009.

[10] Daniel Yergin. Ensuring energy security. *Foreign Affairs*, 85(2):69, 2006.

[11] Kacper Zeromski, Louis Watine, and Jacques Reberol. ENTSOG Yearly Supply Outlook 2022. `https://www.entsog.eu/sites/default/files/2022-07/SO0036-22_Yearly_Supply_Outlook_2022-2023_0.pdf`, 2022. European Network for Transmission Operators for Gas.

[12] Jonathan Stern. Security of european natural gas supplies. *London, the Royal Institute of International Affairs*, 7(5), 2002.

[13] Aad Correlje and Coby Van der Linde. Energy supply security and geopolitics: A european perspective. *Energy policy*, 34(5):532–543, 2006.

[14] Franziska Holz, Philipp M Richter, and Ruud Egging. The role of natural gas in a low-carbon europe: infrastructure and supply security. *The Energy*

Journal, 37(Sustainable Infrastructure Development and Cross-Border Coordination), 2016.

[15] Marco Scotti and Balázs Vedres. Supply security in the european natural gas pipeline network. *Networks in social policy problems*, 2012.

[16] Juan Villada and Yris Olaya. A simulation approach for analysis of short-term security of natural gas supply in colombia. *Energy Policy*, 53:11–26, 2013.

[17] Alena Bašová. Securing gas supply using underground storage. *Social & Economic Revue*, 14(1), 2016.

[18] Benoit Esnault. The need for regulation of gas storage: the case of france. *Energy Policy*, 31(2):167–174, 2003.

[19] International energy agency. `https://www.iea.org/data-and-statistics/data-tables?country=WEOEUR&energy=Natural20gas`, (accessed July 4, 2022).

[20] Aggregated gas storage inventory. https://agsi.gie.eu/, Accessed on 2022-07-04.

[21] International energy agency, gas trade flows. https://www.iea.org/data-and-statistics/data-product/gas-trade-flows, accessed July 4, 2022.

[22] CPLEX User's Manual. Ibm ilog cplex optimization studio. *Version*, 12(1987-2018):1, 1987.

Preparing of a Simulation System for an Examination of Non-Conventional Seating Positions in the Case of a Self-Driving Car

Laszlo Porkolab, Istvan Lakatos

Audi Hungaria Faculty of Automotive Engineering, Department of Road and Rail Vehicles, Szechenyi Istvan University Egyetem ter 1 H-9026 Győr, Hungary
porkolab.laszlo@sze.hu , lakatos@sze.hu

Abstract: In recent years, society has turned it is gaze ever further away from vehicles with conventional drive systems such as the internal combustion engine towards vehicles with electric motors. In the course of these ideas, the concept of autonomous driving is increasingly coming to the forefront and crystallizing as a wish. The article introduces the field of vehicle safety in general and lists possible injury criteria. After that, the current concepts of fully autonomous vehicles will be presented. The aim of the research is to investigate the effect of different rotated seat positions on the injuries of passengers in the event of an accident using a simulation model. The article presents the applied simulation procedure, the simulation results and its validation. The article pointed out that the kinematics of the driver's movement during an accident changes radically as a result of the rotated seat positions, which significantly increases the chance of fatal injuries.

Keywords: Crash test; Occupant safety; Finite element method; Electric vehicles; Autonomous driving

1 Introduction

In recent years, society has turned it is gaze ever further away from vehicles with conventional drive systems such as the internal combustion engine towards new concepts with innovative systems. This includes the development of electric vehicles with electric motors or concept ideas for means of transport with a hydrogen drive. This change is caused by the change in environmental awareness and human curiosity. In the course of these ideas, the concept of autonomous driving is increasingly coming to the forefront and crystallizing as a wish for the future. In addition to a need for protection for the occupants, there is an increase of interest in comfort and economy.

Nowadays, the proportion of accidents caused by incorrect behavior by the vehicle driver still puts any other factors, such as weather and road conditions, in the background as causes for an accident. The introduction of the ESP obligation in 2014 in the European Union for the registration of new vehicles shows, for example, that autonomous systems have a positive influence on road safety. The number of accidents will decrease as the development of autonomous driving continues. The goal should be to achieve zero traffic accidents. Since almost all accidents can now be avoided by the combined application of existing vehicle technology and safe road user behavior, this goal does not seem impossible. [1]

Autonomous driving is divided into five levels. This structuring has established itself worldwide in the automotive industry. The stages of autonomous driving are included in a development process and can also be viewed as a timeline. It is important to note that, while the automatic system only performs its tasks according to predetermined rules, the autonomous systems can learn from those around them and can independently decide how to act. There is no autonomous at level 0. The occupant is the driver of the vehicle and is independently responsible for steering, accelerating and braking. Except for warning systems, no vehicle systems actively intervene in the control. In the next stage, systems are already in place that control either the longitudinal or lateral guidance of the vehicle. Once the driver has selected a particular guide, the system is responsible for the other function. Level 2 is partial autonomous. The driver is able to completely hand over the longitudinal and lateral guidance to the system. The driver is still responsible for monitoring the vehicle and the traffic because the occupant must be able to regain control of the vehicle at any time. With level 3, semi-autonomous driving, it is not necessary for the driver to have to constantly monitor the system. Only in the event of a borderline case does it have to be guaranteed that the vehicle will be taken over after a reasonable handover time. Level 3 is also the current series status of the latest vehicles. At the moment, the development is on the threshold of highly autonomous driving, level 4. The driver is able to completely transfer a specific driving task to the system. These are specific tasks for which the parameters can be narrowed down, which is the main difference to the last level 5, autonomous driving. In the future, there will be no more limits to driverless driving.

As a result of this development, the driver in an autonomous vehicle only finds himself as a passenger. The typical upright and straight sitting position is no longer absolutely necessary and desirable for reasons of comfort. The development of driverless vehicles by the automotive industry is changing driving behavior. The driver becomes a passenger and can perform activities and no longer need to pay attention to controlling the vehicle. Thanks to the freedom gained, it is no longer necessary for the driver to remain in an upright sitting position facing straight ahead. An important research direction can be the influence of a rotating seat in the first row of seats in a car on future restraint system is shown.

2 Vehicle Occupant Safety

Vehicle occupant safety includes active and passive safety. A distinction is made here, which is structured as follows: measures to limit the consequences of an accident (active safety) and systems to reduce the consequences of an accident (passive safety). The active safety systems intervene to support the driving operation. The detection of the driving condition is recorded and evaluated with the help of various sensors. In the event of deviations from defined limit values, the driving condition is intervened.

For example, the electronic stability program (ESP), which has been mandatory for all new vehicle registrations since 2014, detects acute dangerous situations, such as rotations around the vertical vehicle center axis and engages through a specially calculated braking of individual wheels. Another system is the anti-lock braking system (ABS). This may prevent the vehicle from swerving when braking. In order to protect the occupants of a vehicle from serious or even fatal injuries in the event of an accident, the passive safety systems must come into effect. The vital parts of the body such as the head and chest are particularly safe. The consideration of these areas takes place in-depth. The combination of active and passive safety, as well as the consideration of the entire course of the accident from the origin of the accident to the rescue service, lead to integral safety. The aim is to optimize the protection potential of all road users.

The governments of different countries have taken several measures to improve road safety for vehicles. This is based on the UN / ECE regulations that were established in 1958 by the United Nations Economic Commission for Europe (UNECE). Nowadays, the regulations for North America are laid down in the Federal Motor Vehicle Safety Standards, or FMVSS for short, whereas in Europe the ECE regulations (Economic Commission for Europe) form the basis for road safety regulations. An essential difference between these regulations are the unbelted load cases in the United States. Other countries have their own laws. [2]

The statutory regulations stipulate minimum requirements for individual components and assemblies that must be met. If the requirements of the law are not met, there is no type approval for the respective vehicle. Since the statutory provisions only reflect a fraction of the total number of accidents, additional consumer protection load cases have been developed by independent private institutes. The most important organizations for the continuous development of consumer protection in motor vehicles include the NCAP institutes (New Car Assessment Program) and the IIHS (Insurance Institute for Highway Safety). These crash tests are conducted by NCAP institutes such as C-NCAP (China), Euro-NCAP (Europe), US-NCAP (United States) and the IIHS (Insurance Institute for Highway Safety (North America). The NCAP programs are adapted to the respective regions. One example is the different requirements in the event of a side crash in the USA, because the proportion of large vehicles such as SUVs

(Smart Utility Vehicles) is significantly larger there. Another difference between the individual NCAP institutes is the different evaluation method. The aim of the legislation and consumer protection regulations is to make an objective safety assessment possible in the automotive industry. The automobile manufacturers are not able to do this obliged to pass the consumer protection load cases positively, but the results are freely available he customer the opportunity to compare vehicles from different manufacturers with regard to their safety.

A crash is a very sharp and abrupt deceleration of the vehicle. With the law of conservation of momentum, the occupant tries to move on with the speed vector after an impact. The use of different restraint systems includes the task of reducing the relative speed that prevails between the occupant and the vehicle so that the impact of the occupant on the vehicle interior structures does not have fatal consequences and a reduction in the severity of injury is achieved. A reduction in the impact speed is achieved by holding the passengers in the seat. The seat ramps are coupled to the seat to prevent the submarining effect, which reduces the pelvis from sliding forward and thus makes it more difficult for the occupant to slip under the lap belt. The restraint effects of the airbag and the seat belt are not impaired. The headrests belonging to the seat prevent or reduce the overstretching of the cervical spine and the associated serious injuries in extreme cases and mainly in the event of a rear impact. The basis for occupant protection systems is generally the body of each vehicle, which is not explicitly assigned to the occupant protection system. The body has two crucial tasks that are initially mutually exclusive. On the one hand, the external structure of the vehicle is responsible for providing a sufficiently high level of rigidity for the passenger cell so that the survival space for passengers is guaranteed. On the other hand, it is necessary that the deformation elements of the body are sufficiently deformable in order to convert sufficient kinetic energy into deformation energy.

2.1 Injury Criteria

Classification of injuries requires a comprehensive evaluation principle. The AIS (Abbreviated Injury Scale) is most often used in accidents, is an anatomically based global severity scoring system. This scale basically expresses how life-threatening the given injury is. Its very first version was published in 1969, but since then there have been many updates, the most recent in 2015. Each injury is characterized by three parameters on the scale, these are type, location and severity. The scale classifies the types of injuries into six groups, these are whole area, vessels, nerves, organs, skeletal and loss of consciousness. The scale classifies the location of injuries into a total of nine groups, these are head, face, neck, thorax, abdomen, spine, upper extremity, lower extremity and unspecified. The scale classifies the severity of injuries into a total of seven groups, these are minor, moderate, serious, severe, critical, fatal and not further specified. [3]

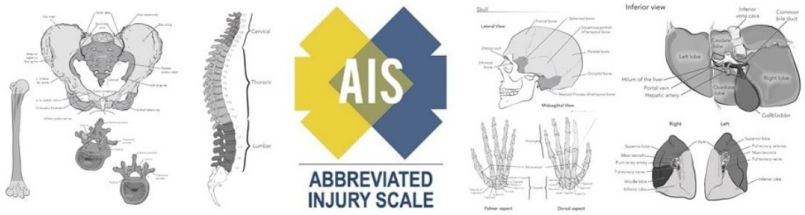


Figure 1
Abbreviated Injury Scale [3]

During crash test analysis the most important measured values in the head area include the translational accelerations, which are measured in the center of gravity of the head. From these, the head acceleration value a3ms head and the HIC (Head Injury Criterion) are calculated. The head acceleration value a3ms, which is specified over a period of t=3 ms, represents the greatest acceleration of the head. For legal requirements, such as according to FMVSS208 (United States), a value below 80g must be met. In order to achieve a maximum number of points at the Euro-NCAP, this must not exceed the value of 72 g. Acceleration values in the x, y, and z directions are displayed as a_x, a_y, and a_z. The limit value for head injuries due to acceleration is defined for a time interval of 15 ms, which represents a hard impact, and a time interval of 36 ms, which describes a softer head impact. For example, in FMVSS208 (United States) load cases for frontal impact, a HIC of 700 is given, while the values required to reach the maximum Euro-NCAP frontal impact score must be below 500.

Another important area of injury is the neck. The Nij (Normalized Neck Injury Criterion) is used to assess cervical vertebrae injuries, taking into account occupant size. The causes of injuries to the cervical spine lie in the acting axial tensile and compressive forces and the bending moments around the transverse axis at the transition from head to neck, which is why these values are used to calculate the injury criterion. To determine the loads acting on the chest, the chest compression is determined in addition to the chest acceleration a3ms. In general, according to FMVSS208 (United States), a relative compression path between the sternum and the spine of 50.8 mm must not be exceeded in the case of contact between surfaces of any kind. This would cause serious internal and external injuries in the area of the upper thorax. In the event of contact with the airbag system, the upper limit is 76.2 mm based on the assumption that force is applied over a larger area. When using belt systems, the lower limit value must be observed.

Knee injuries are increasingly occurring due to very large pelvic displacement values. This is the case in particular with unbelted occupants. The cause can also be found in connection with the so-called submarining effect, in which the vehicle occupant can slip under the seat belt. In addition, foot space intrusions restrict foot movement, which is why pressure and bending stresses arise in the lower leg,

which are associated with a high risk of fractures and ligament ruptures. Due to the different forces acting between the thigh and tibia, a displacement in the knee can occur. Therefore, this displacement, also known as the kneeslider effect, between the upper and lower leg was defined as a protection criterion for the knee joint. Since bending moments occur in addition to compression forces when impacting the instrument panel, the tibia index is used to assess the risk of injury to the tibia.

3 Current Concepts for Fully Autonomous Vehicles

Today the automotive industry is making great strides towards driverless driving. In this chapter, four concepts from the automotive industry are shown in order to show the variety of seating positions of the visions. The first model to be presented is the F015 concept study from Daimler AG. The futuristic design combines many different new ideas. On the one hand, the vehicle structure is fundamentally different compared to today's vehicle models. The lack of a B-pillar means that the interior of the vehicle can be given completely new functions. In *Figure 2* it can be seen that this vehicle concept is equipped with four individual swiveling seats. The angle of rotation that can be achieved with these vehicle seats is not clearly visible, but this Figure shows a face-to-face constellation of the four seats.

Figure 2
Vehicle concept (left) and interior concept (right) of the F015 from Daimler AG [4]

With its concept study 360c, the Swedish car manufacturer Volvo offers another idea for driverless driving (see *Figure 3*). However, this concept offers several variants compared to the concept studies from Daimler AG. In general, the face-to-face sitting position is given for the interaction between occupants. However, there is also the variant in which the seat back is set almost horizontally and the occupant is in a lying position. The concept shows that sleeping shouldn't be a problem in general, as it will be a fully autonomous vehicle.

Figure 3
Vehicle concept (left) and interior concept (right) of the 360c from Volvo AG [5]

Most of the concept studies presented are image advertising. However, these concepts offer insights into possible future design ideas and ideas of self-driving vehicles. None of the concept ideas mentioned go into detail on adjustments to restraint systems to protect passengers. The research phase is on to change the impact boundary conditions and the behavior of current restraint systems and,consequently, sensible solutions. In addition, protection systems and new concepts for protection systems can only be designed once the vehicle structures have been coordinated.

There are some taxis on level 4 without driver in use one of them is the Cruise AV. The Cruise AV is a Chevy Bolt-based autonomous vehicle; the first generation (G1) were modified by Cruise in San Francisco while the subsequent second and third generations (G2, G3) are manufactured at the Orion Township assembly plant in Michigan. The Cruise AVs feature drive control algorithms and artificial intelligence created by Cruise. The Cruise AV uses Lidar, radar, and camera sensors. In September 2021, Honda started testing program toward launch of Level 4 mobility service Business in Japan, using the G3 Cruise AV.

Figure 4
Cruise AV self-driving taxi [6]

Another important company developing driverless taxis is a Waymo. Waymo is an American autonomous driving technology development company. It is a subsidiary of Alphabet Inc, the parent company of Google. Waymo operates a commercial self-driving taxi service. In October 2020, the company expanded the service to the public, and it was the only self-driving commercial service that operates without safety backup drivers in the vehicle at that time.

Figure 5
Waymo self-driving taxi [7]

4 Description of the Simulation System

4.1 The Algorithm of the Simulation System

Due to the rapid technical development in recent years, it has become possible to solve static and dynamic problems mechanically with the help of computer simulations. This change from physical, tangible product creation to virtual product creation is becoming more and more common in industry. Computer Aided Engineering (CAE) is making a significant contribution to this trend. It offers the possibility of computer-aided modeling (CAD) through computer-aided execution and evaluation (CAT) to computer-aided manufacturing (CAM). In order for a component to be simulated and calculated, it is necessary to transfer the geometry defined in the design file to an FEM (Finite Element Method) model.

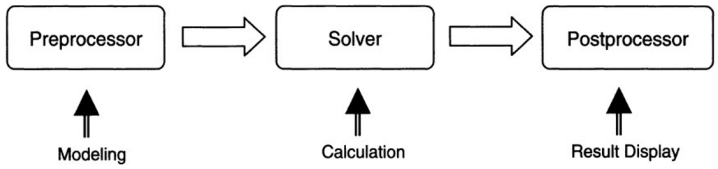


Figure 6
Simulation steps [8]

In order to carry out this activity and to model the necessary adjustments, the ANSA preprocessor from BETA CAE Systems used in this work. After the boundary conditions and the model have been transferred, the solver is able to determine and solve the differential equations and save them in an output file for each time step. The LS-Dyna solver from Livermore Software Technology Corporation is used for the simulations set up to calculate the FE model. By using

the postprocessor Animator 4 from GNS mbH and in combination with ANSA META, the results in this work are evaluated and made available graphically. ANSA META is a postprocessor developed explicitly for ANSA. Analysis results include deformations and movements over time, as well as stresses and natural frequencies can be displayed.

The finite element method describes a mathematical-numerical method for solving physical problems. Areas of application include structural mechanics, heat transfer, and fluid mechanics. In order to be able to determine the stresses and deformations of the systems to be examined, the structure must be divided into a large number of finite elements, the so-called finite elements. Depending on the modeling, suitable elements are used. These can take on one, two or three dimensional forms. Solid elements (tetrahedron or hexahedron elements) are required for a three-dimensional continuum. [8]

The use of special elements such as the shell elements for the representation of "thin-walled" components, for example sheet metal structures, can be useful and are assigned to the two-dimensional forms. In addition, there are also bar elements or spring and damper elements. The combination of different element types is possible and necessary in the field of multi-body simulation. The vertices that form the finite element are called nodes. The node movement in three-dimensional space and the associated deformation of the bodies are calculated for each individual node. Partial differential equations of the 2nd order are set up and solved. Acceleration, velocity and position are calculated for each time step. A separate stiffness matrix, damping matrix and mass matrix is set up for each individual element. This matrix depends on the batch function and the material used. The shape functions have an important role. By using these functions, it is possible to draw conclusions about stresses and strains from the calculated nodal displacements. Here, the relationship between the shifts in an element becomes apparent. The higher the degree of the interpolation function, as the shape function is also called, the more exact the result.

There is an implicit and an explicit time step method for the simulation calculations. The implicit method does not fix a time step. Here, not only the known variables at the beginning of a time step are included, but also those sought for the following time step. The equations of motion are calculated iteratively by calculating back from t + x to t. This calculates new values from the old and new values. [8]

The explicit time discretization uses time steps that have a given length. The time step size depends on the calculation accuracy and the calculation time because narrowing the time step width increases the accuracy of the results, but the calculation time increases in return. The accuracy increases, but only up to a certain limit. Narrowing the time step then does not lead to any further improvement. In general, the time steps should be chosen so that they are smaller than the meshing of the elements. An advantage of the explicit method compared

to the implicit method is the lower performance requirement of the computer systems. Likewise, non-linear phenomena (e.g. contacts between two components) can be calculated more easily. The advantage stems from the fact that no systems of equations have to be solved. With the advantage also comes a disadvantage. The explicit method can only be used for relatively short computing times. For large time steps, the implicit time integration provides more accurate results, which are, therefore, very stable. For this reason, the implicit method is used for linear or weakly non-linear problems.

This research is a crash simulation with occupant models. In crash simulations, non-linear dynamic processes occur with a short duration of action. Accordingly, the explicit time discretization is used. In the event of vehicle collisions, the stresses that occur are not used as the main source of information since the plastic strains and the deformation energy absorbed (internal energy) are of greater importance. In the course of deformation, elements can become "smaller". This influences the critical time step since it becomes smaller and the computational effort increases. However, the critical time step can be almost completely compensated with the help of so-called "mass scaling". The basis is the scaling of the density and thus the mass. As a result, the speed of sound, which is dependent on the density, decreases and the time step remains the same. This procedure creates certain calculation errors that are negligible within a certain framework.

Crash simulations are multi-body systems. There are a large number of different components and bodies with which the calculation is started. The contacts form the basis for the optimum functioning of the simulation. It is necessary that each individual component has a self-contact. This makes sense so that a component can come into contact with itself and the kinematics can be reflected realistically. The different components are only able to influence each other if contacts between them are defined. Otherwise, these components can be moved freely in space and penetrations are visible. With these contacts, each node of the contact surface (master) is checked in each calculation step to determine whether it is in contact with an element of the target surface (slave). In order to solve convergence problems, the penalty formulation is used in crash simulations, among other things. This algorithm allows for a certain overlapping of the contact surfaces. In return, a spring force that depends on the contact stiffness is applied to the nodes that come into contact. This power is called the penalty power. Depending on the type of stress and the problem, different types of contact are possible. It is possible to define several components of a simulation at the same time with a further number of components with one contact. This is necessary and useful for complex occupant simulations.

4.2 Loadcase Definition

When analyzing and evaluating the results, only the restraint systems of the respective load case and the dummies are evaluated with regard to certain criteria. Furthermore, there is no evaluation of unstrapped load cases. In addition, there is no provision for an evaluation of the passenger or the occupants in the second row of seats. So this research only deals with the driver's seat side. In order to best represent the influence of swivel seats, care was taken to select the critical loadcase.

There are two types of frontal impact: impact with a rigid wall and deformable barrier with an offset. The crash with a rigid wall is consequently the impact with the maximum forces and deformations, because the wall does not allow any deformation and accordingly no energy dissipation and the energy flow is completely guided through the vehicle. In contrast, the impact with a deformable barrier converts part of the kinetic energy that occurs into deformation energy and thereby lowers the part of the kinetic energy that has to be converted into deformation energy by the vehicle body. A uniform, comparable evaluation with simultaneous high stress for the occupant is only guaranteed by the load case with the impact on a rigid wall, since the positions of the dummy and the seat are not only linked to a uniform pivot point, but also to a translation. The position of the occupant would move further and further into the interior of the body and thus strong local deformation would escape. For this reason, only the loadcase with impact on a rigid wall is considered in this work. [9]

4.3 Seat Position Definition

To investigate the influence of swivel seats on the restraint systems, the seat positions to be analyzed must first be narrowed down. Already in Chapter 3 when introducing existing vehicle concept studies, the maximum angle of rotation used is opposite to the normal direction of travel. *Figure 7* shows a matrix according to which the different swivel seat position variants are named in this work.

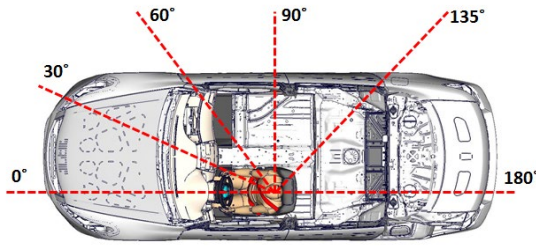


Figure 7
Representation of the examined angles of rotation (Source: Author's plot)

The area examined extends from 0˚, which corresponds to the current driver's seat position, to the 180˚ position, in which the driver's view is directed towards the rear. In order to limit the number of variants that are examined, the angle steps are limited. As part of this thesis, the 0˚, 30˚, 60˚, 90˚, 135˚, 180˚ positions of the seat are examined in more detail and examinations are carried out on the basis of these variants.

The selected rotation points can be seen in *Figure 8*. The verification takes place by building up the various angles of rotation for each point. The result shows that rotating the driver's seat without adapting the vehicle interior is impossible. In addition, it can be seen that a single rotation around a selected pivot point is not sufficient to control the selected rotation angle, because depending on various degrees of angle, the seat overlaps with the vehicle environment, in particular with the steering wheel and the B-pillar. An adjustment of the position of the steering wheel and the B-pillar are impossible due to large geometrical interventions.

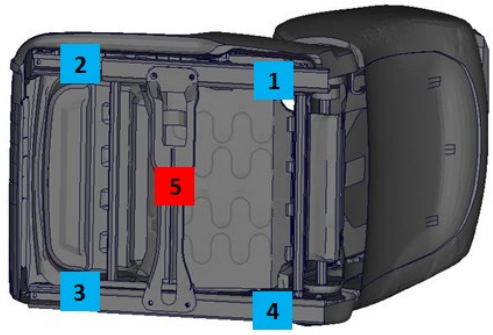


Figure 8
The examined rotation points of the seat console (Source: Author's plot)

To remedy this, two translations are introduced in combination with the rotation. The translations of the seat and the dummy on it ensures a sufficient distance from the steering wheel and the B-pillar. The fifth rotation point is used as the starting point for the rotations and translations. The starting point for determining the appropriate translations is the 90˚ sitting position. The head position is controlled accordingly with the aim of an exact impact on the driver's airbag. Starting from this 90˚ base position, the translations of the other angular positions are determined using a linear dependency. The proportion calculation is demonstrated using an example. The 90˚ position represents 100%, so 30˚ are considered a third of the base. Consequently only 30% of the translations that are used in the 90˚ position are to be applied to the 30˚ position. [10]

4.4 Adaptation of the Vehicle Model

For the following investigations, the Honda Accord model is adapted to the loadcases of the respective crash type. For a front impact, a rigid wall is placed in front of the vehicle model and the Honda Accord is hit at a speed of 50 km/h. It should be noted that the wall and the vehicle have a 100% overlap.

The simulation model is a conventional vehicle. Accordingly, measures are to be determined that lead to the elimination of the lack of space inside the vehicle. In order to generate the space taken up for the rotation of the seat, the front passenger seat is removed at the beginning. The second step is to remove the entire center console of the vehicle. This is what makes it possible for the first time to apply a rotation to the driver's seat without penetrations. In Chapter 3, no center consoles can be seen in the concept studies presented in the area of the driver and front passenger. The adaptation of the interior is marked in *Figure 9*. [11]

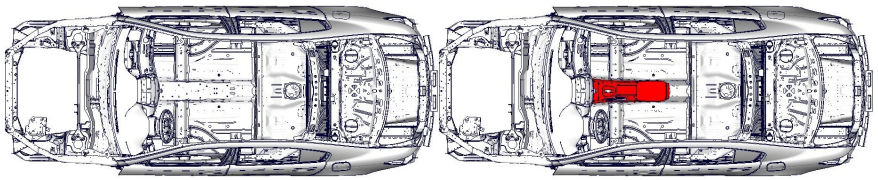


Figure 9
Base model (left) and adaptation (right) (Source: Author's plot)

In addition, the underbody of the vehicle model must be adapted to the boundary conditions so that it is possible to turn the driver's seat into the respective rotational positions. In order to gain space-related advantages, the tunnel to which the center console is attached is flattened. Electric vehicles are often associated with driverless driving. In the case of electrically operated vehicles, the exhaust gas system is omitted due to the lack of exhaust gases, as is the tunnel designed for this purpose, if this is not used for cooling. In addition, in current vehicle models, the battery, energy storage, is attached below the underbody. *Figure 10* shows the geometric changes made to the sub-floor. [12]

Figure 10
Geometric adaptation of the sub-floor (red: base; blue: adaptation) (Source: Author's plot)

In addition to the points mentioned above, it is essential to adjust the seat. In every vehicle, the seat is attached to the so-called seat cross members, which give the seat the necessary stability with regard to various accident scenarios. As a rule, current vehicle models have two seat cross members, but the Honda Accord model is one of the vehicle models that only have a single seat cross member. As part of this research, fastening the seat console to the seat cross member by rotating the seat is neither sensible nor possible, as massive lever arms occur and would influence the calculation. The task of the seat cross member is to connect the seat to the vehicle body via its seat console, thereby ensuring the restraint effect and the flow of forces. For this research, the seat is connected to the underbody of the vehicle for every rotation and translation. Accordingly, a separate mounting base is created individually for each angle of rotation in the interval from 0° to 180°. This creates an optimal anchoring of the seat console on the underbody in the simulation model. [13]

An essential change to the present model is made in the attachment of the belt. The seatbelt is usually anchored in the B-pillar due to the large amount of force exerted, but the driver's seat cannot be rotated with the belt attachment as the belt would tighten the occupant's neck. For this reason, the seatbelt holder is integrated within the seat. A change in the direction of force flow and an additional load on the seat are to be expected. The changes are shown in *Figure 11*. [14]

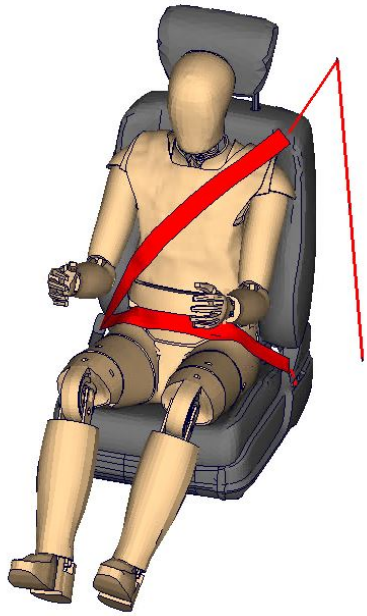


Figure 11
Geometric adaptation of the seat belt (Source: Author's plot)

5 Results of the Simulations

In the following chapter, an overview of the results of all examined angles of rotation during a frontal crash is presented. Only the driver side position was examined during the simulations. On the basis of this preliminary investigation, certain trends can be established for certain properties of the angle of rotation. In the further course of the process, the resulting abnormalities are substantiated by a detailed examination of the dummy kinematics and the consumer protection criteria. This is necessary in order to make a selection from the simulations used for further investigations. Critical load cases are then selected on this basis. The comparison of the simulations is made in connection with the most important evaluation criteria. [15]

In order to provide an overview of the examined seating positions, the measuring dummies are divided into the load cases. In order to get a first impression of the determined results, the evaluation takes place with the help of the limit values of the most important Euro NCAP criteria. For the angle ranges from 0° to 90°, the driver airbag is ignited with an ignition time of 14 ms. On the other hand, there is no ignition of the driver airbag in the area of the rotation angle from 135° to 180°, since the effective range of the airbag is significantly restricted by the seat backrest and a large part of the protective effect is lost. The results of the sitting positions for the angles of rotation from 0° to 180° are shown on *Table 1* and *Table 2*. The following tables show the results of the simulations with the adopted model.

Table 1
Results overview (yellow = maximum value)

Criteria	Limit value	Unit	Simulation 0°	Simulation 30°	Simulation 60°
Head (HIC15)	700	[-]	487	1159	964
Head (a3ms)	80	[g]	55,1	91,3	93,1
Head (BrIC)	1,05	[-]	0,57	0,93	0,95
Neck (Nij)	0,85	[-]	0,43	0,59	0,48
Chest (Compression)	60	[mm]	41,4	50,6	54,9
Abdomen (Compression)	88	[mm]	68,1	113,4	103,5
Femur (Force)	7,56	[kN]	6,88	4,39	6,11

Table 2
Results overview (yellow = maximum value)

Criteria	Limit value	Unit	Simulation 90°	Simulation 135°	Simulation 180°
Head (HIC15)	700	[-]	833	803	731
Head (a3ms)	80	[g]	83,3	75,3	72,9
Head (BrIC)	1,05	[-]	1,74	1,43	1,18
Neck (Nij)	0,85	[-]	0,57	0,63	0,61
Chest (Compression)	60	[mm]	48,3	46,1	56,7
Abdomen (Compression)	88	[mm]	91,4	78,6	71,4
Femur (Force)	7,56	[kN]	6,32	8,17	7,26

6 Validation of the Simulations

It is absolutely necessary to use a real test to validate the simulations. The simulations were based on the 2008 model of Honda Accord, therefore, requisite to use the real test of this model as a basis for validation. For this front impact test, a rigid wall is placed in front of the vehicle model and the car hit at a speed of 50 km/h with 100% overlap.

The tested model of Honda Accord Euro was introduced in Australia during 2008. Dual front airbags, side airbags and head-protecting side curtains are standard equipment. Antilock brakes (ABS), electronic brake distribution (EBD) and electronic stability control (ESC) are also standard. Intelligent seatbelt reminders are fitted to all seats. The front seatbelt buckles are mounted on the seats and the upper anchorages are adjustable. These features improve the fit of the seatbelt. Pretensioners are fitted to the front seatbelts to reduce slack in the event of a crash. A three point seatbelt is fitted to the centre rear seat. This provides better protection than a two point seatbelt. [15]

The Accord Euro scored 14.47 out of 16 in the frontal crash test (5 Stars). The passenger compartment held its shape well. There was a slight risk of serious chest injury for the passenger and a slight risk of serious lower leg injury for the driver and passenger. Body region scores out of 4 points each: Head/neck 4pts, chest 3.45pts, upper legs 4pts, lower legs 3.02pts. The accelerator pedal moved rearwards by 41 mm. The steering wheel hub moved 20 mm forward, 13 mm upward and 7 mm sideways. The front A-pillar moved 16mm rearwards. All doors remained closed during the crash. After the crash all doors could be opened with

normal effort. The airbag cushioned the head of the driver and contact was stable. There were no knee hazards. The passenger's head was cushioned by the airbag.

Maximum points were scored for protection of the 3 year infant based on the dummy result from the impact tests. The passenger airbag can be disabled to allow a rearward-facing child restraint to be used in that position. However, information presented to the driver regarding the status of the airbag is not sufficiently clear. A warning label clearly warns of the danger of using a rearward facing child seat in the passenger seat without first disabling the airbag, but the information is not available in all European languages. The presence of ISOFIX anchorages in the rear outboard seats is not clearly marked.

The bumper scored maximum points for the protection offered to pedestrians' legs in the area rated by Euro NCAP. However, additional tests showed that some areas beyond the central zone provided poor protection. Most of the zones tested in the area of the bonnet where a child's head might strike also scored maximum points. However, the front edge of the bonnet was rated as poor.

Figure 12
Honda Accord 2008 EUNCAP Crash Test [16]

Table 3 shows the results of the base model and the modified model in comparison with the crash test. The base model is validated, it can be seen that all differences in the simulation results remain within the 10% limit. The situation is different for the fitted model, due to the attachment of the belt in the seat and the associated reduced restraint effect. This results in a significantly changed kinematics of the seat and the dummy and the results also show a partially larger difference compared to the crash test. There is a bigger difference in the values of the neck and chest, the main reason for this is the change of the seatbelt. The fitted model should be used for comparison with rotated seat variants, all modifications are summarized in Chapter 4.4.

Table 3
Results of the base model and the modified model in comparison with the crash test

Criteria	Unit	Limit value	Crash test	Simulation 0° Base model without any modifications	Simulation 0° Fitted model with all modifications
Head (HIC15)	[-]	700	416	437 (+5%)	487 (+17%)
Head (a3ms)	[g]	80	47,4	49,4 (+4%)	55,1 (+16%)
Neck (BrIC)	[-]	1,05	0,39	0,42 (+8%)	0,57 (+46%)
Neck (Nij)	[-]	0,85	0,64	0,63 (-2%)	0,43 (-32%)
Chest (Compression)	[mm]	60	32,8	34,3 (+5%)	41,4 (+26%)
Abdomen (Compression)	[mm]	88	72,9	70,7 (-3%)	68,1 (-7%)
Femur (Force)	[kN]	7,56	6,03	6,25 (+4%)	6,88 (+14%)

Conclusion

The aim of the research was to create a computer simulation model suitable for testing the passive passenger safety of a fully self-driving vehicle. The computer model was validated with measurements and transformed in such a way that it is suitable for testing the desired seating positions. We examined to what extent the seating position of the passengers affects their injuries in the event of an accident. During the tests, the results of the injuries that occurred in the case of seat positions turned by 30°/60°/90°/135°/180° we compared with the case of the driver's seat in the normal basic position. We pointed out that the kinematics of the driver's movement during an accident changes radically as a result of modified seating positions. We have proven a large reduction in the effectiveness of traditional passive protection systems, and the chance of fatal injuries increases significantly in the case of rotated seating positions. This applies primarily to head and neck injuries, but the impact value of the chest and the extent of leg injuries also become more serious.

That is why it is necessary to expand the passive protection systems known today. Our goal is to create a system that provides effective protection even in the case of rotated seating positions. We examine the possibilities of further developing the driver's seat from the point of view of passive vehicle safety, what kind of protection it could provide for the fully self-driving car of the future. The seat must be able to coordinate the movement of passengers in the event of an

accident, even in rotated seating positions, and must be involved in preventing serious injuries. We build a model to test and verify this.

In addition to the above, we would like to investigate new possibilities for the use of airbags in self-driving vehicles. In a vehicle without a steering wheel, the airbag for driver protection can be imagined in a radically new position, size and shape. We determine which direction of change affects the driver's injuries in the event of an accident.

References

[1] European Commission - Questions and Answers, "Road safety statistics 2020: what do they mean a numbers?", 2020

[2] A. Kullgren, A. Lie, and C. Tingvall, "Comparison between Euro ncap test results and real-world crash data", Traffic Injury Prevention, vol. 11, pp. 587-593, 2010

[3] E. Petrucelli, J. D. States, L.N. Hames, "The abbreviated injury scale: Evolution, usage and future adaptability", Accident Analysis & Prevention, vol 13, pp. 29-35, 1981

[4] Daimler, "Mecedes-Benz F015 Luxury in Motion", Daimler AG, 2017

[5] Volvo, "Volvo cars in the future", Volvo Car Corporation, 2018

[6] Criuse, "G3 Cruise AV based on the Chevrolet Bolt", Cruise LLC, 2020

[7] Waymo, "Waymo One", Waymo LLC, 2020

[8] P. Fröhlich, "FEM application practice. Introduction to Finite Element Analysis", bilingual edition german/english, Studium Technik, 2005

[9] J. M. Nolan, A. K. Lund, „Frontal offset deformable barrier crash testing and its effect on vehicle stiffness", Insurance Institute for highway safety, 2001

[10] M. S. Samad, M. K. Mohd Nor, "Optimization of vehicle pulse index parameters based on validated vehicle-occupant finite element model", International Journal of Crashworthiness, vol. 28, pp. 217–223, 2023

[11] A. Leledakis, M. Lindman, J. Östh, "A method for predicting crash configurations using counterfactual simulations and real-world data", Accident Analysis and Prevention, 2021

[12] J. O. Hallquist, "LS-DYNA R Theory MManual", Livermore Software Technology Corporation, Livermore, California, 2006

[13] M. D. Marin, J. A. Alonso, "Child safety in autonomous vehicles: living room layout", Dyna Ingenieria e Industria, vol. 97, pp. 30–34, 2022

[14] K. Golowko, V. Zimmermann, and D. Zimmer, "Automated driving influences on the restraint system", (in German) ATZ-Automobiltechnische Zeitschrift, vol. 119, pp. 26–33, 2017

[15] W. M. Choi, H. Y. Jeong, "Design methodology to reduce the chest deflection in usncap and euroncap tests", International Journal of Automotive Technology, vol. 13, 2012

[16] ANCAP, "Honda Accord 2008 crash testing for safety", New Car Assessment Program, Australia, 2009

The Implications of Electric Scooters as a New Technology Artifact in Urban Transportation

Dorottya Szemere, András Nemeslaki

Budapest University of Technology and Economics, Műegyetem rkp. 3, 1111 Budapest, Hungary, szemere.dorottya@bme.hu; nemeslaki.andras@gtk.bme.hu

Abstract: E-scooters are becoming an increasingly important component of urban transportation networks, providing a sustainable alternative for first and last-mile travel. However, if they are not managed and regulated carefully, they can become a risk to users and cities in general. In this paper, we studied the emergence of this innovative, platform-based, and shareable e-scooter, as a possible 'greener' transport alternative, from a sustainability and environmental awareness perspective. The most critical question is legislation. Urban planners need to implement laws limiting maximum speeds, requiring the use of e-scooter infrastructure, and providing dedicated parking, as well as limiting the number of licensed operators. In our exploratory research, we conducted three focus group discussions to explore the views of non-users and then built the nomological network from the key phenomenons they identified. The participants are confused about micro-mobility solutions, according to the findings, due to a lack of clarity on regulatory problems and poor infrastructure. As a result, we propose that policies should be co-developed in cooperation with various stakeholders to enable antagonistic conversation about alternative policy actions by evaluating their effectiveness in terms of behavioral change as well as their implementability.

Keywords: sustainability; electric scooter; nomological network

1 Introduction

E-scooters are a relatively new topic in transportation research. This new vehicle type originally arrived in the United States, California in 2017 and is now available all over the world. After car sharing and bike sharing, most e-scooters are organized in sharing networks, making them the newest invention of shared mobility. These schemes have become essential to most large city transportation systems in North America, Asia, Australia, and Europe (Weschke – Oostendorp & Hardinghaus). There are approximately 20 million users in Europe alone, and the adoption rate of e-scooter sharing is four times that of bike sharing (Latinopoulos – Patrier & Sivakumar).

2 Framework

To date, the scientific basis for the argument about the use and impact of e-scooters is limited. There is a lack of data from numerous studies, and it is unknown to what extent the data has been systematically obtained, examined, and interpreted. On this premise, it is unclear whether and how e-scooters can genuinely contribute to more sustainable mobility. Even though we can accurately conclude that the meaning of technology is socially constructed. Our research requires an understanding of technology as well as the organizational and individual difficulties associated with its application (Benbasat – Zmud, 2003).

The most popular statements in previous research about e-scooter are the following:

- E-scooters are seen as an important part of the evolution of cities transportation systems since they 'may be used for a considerable proportion of in-town travel' (Schellong et al., 2019, p. 2) and can be coupled with other modes, especially public transportation.
- E-scooters are promoted as a low-cost, socially inclusive means of transportation as part of micro-mobility (Semenov, 2017).
- E-scooters are marketed as a 'green solution for urban traffic' (Kopplin et al., 2021, p. 2).
- E-scooters are mostly seen as 'fun objects' (Kopplin et al., 2021, p. 1), but not as a serious mode of transportation.

The above examples of statements taken from various research studies illustrate that the conceptual network that gives meaning to electric scooters is still in its emerging stage. This paper aims to conceptualize the electric scooter as an artifact of the application of electric micro-mobility to support or complement other urban transport methods embedded within government policies that are embedded within the community's attitude towards sustainability.

3 Problems and Research Questions

E-scooters are very popular nowadays, and the story of electric scooters as a vehicle dates back more than a century. The first motorized scooter, called Autoped, was launched in 1915 and quickly became popular in the United States (Manskie, 2019). Autopeds, like their modern analogs, was accused of being a toy for the aristocratic people. Furthermore, there is evidence that the method was used for traveling and leisure uses, as well as for enterprises such as postal services, which used it to distribute letters. The main constraints limiting their commercial success were their high cost compared to bicycles and insufficient comfort compared to motorbikes. In more recent years, Go-Ped released the first manufactured motorized scooter since 1915. During the previous decade, with the invention of lithium-ion batteries

and technological advances in electric motors, the Go-Ped design began to be changed into the first e-scooter prototypes. E-scooters are significantly less expensive than their gas-powered equivalents. However, it appears that the important step in their popularity was the greater awareness that followed their introduction with the expansion of bike-sharing programs. Even though shared e-scooters have only been on the market for four years, they are quickly becoming an important factor in urban areas, changing our user behavior.

The emergence of e-scooters raises the question of whether they can become a disruptive niche innovation. According to Geels et al. (2017), any transformation of an existing transportation system necessitates changes in technologies, infrastructures, organizations, markets, legislation, and user practices. This new way of electric micro-mobility rivals cars in terms of speed, safety, and affective values. Because any innovation is generally faced with resistance in the prevailing system, its introduction must be carefully considered because it disrupts established government transportation regulations (Gössling – Cohen, 2014). While academic discussions of restrictions to e-scooter systems have focused on a lack of charging infrastructure, reduced subjective safety for other traffic participants, adverse weather conditions, or a limited capacity to transport baggage (Hardt – Bogenberger, 2019: 155), global newspaper reports (e.g., Guardian, 2019) would appear to suggest that public opinion is the most significant challenge for this mode of transportation.

As a part of this larger global trend, Hungary has seen the emergence of standing, shared, or private, rechargeable electric scooters and other micro-mobility devices. The public has replied to e-scooters with both enthusiasm and skepticism, as communities have struggled with unanticipated consequences such as irresponsible riding, cluttering, and vandalism. This new form of transportation poses challenges for all transport actors. In consideration of the possibilities and limitations of e-scooter technology, our question is whether this type of micro-mobility can launch a larger urban transportation system change. In this study, we present the results of three focus group discussions aimed at gathering the views of non-users on regulatory issues related to electric scooters and providing policy stakeholders with action-oriented guidance.

4 Research Methodology

At this early stage of our research, we focused on how this new form of electric micro-mobility is received by those stakeholders who do not use electric scooters. In this article, we present their attitudes toward e-scooters, government policies, alternative transportation modes, and sustainability. Because the electric scooter is a relatively new phenomenon in urban transportation about which we don't know much and the conceptual framework around this term has not yet been stabilized,

we used focus group interviews as a possible qualitative research method to build the nomological network to illustrate the broad public opinion and how electric scooters could be integrated into urban transportation infrastructure. This methodology allowed the researchers to examine how the participants interacted when discussing their ideas and perspectives (Liamputtong, 2012). This method of inquiry allows for the study of small groups of people who have firsthand knowledge of the subject under examination (Blumer, 1969).

According to Wilkinson (2004), the assemblement and interpretation of human experience are affected by others. Participants in the focus group discussions brought their unique stories to the meeting and worked to develop a common understanding of them (Wilkinson, 2004). As a result, we were able to observe the members in the focus group co-constructing meaning. The use of focus group methodology in this study provided a mechanism for us to improve our knowledge of the subjective meaning that participants ascribed to their behaviors and circumstances (Flick, 2006).

Respondents were selected using convenience sampling and the snowball method. The first group of participants was chosen using the contact network of the FIEK organization of the Budapest University of Technology and Economics. The FIEK is responsible for managing the relations between the university and large companies, and therefore The FIEK is responsible for managing the relations between the university and large companies, and therefore the first focus was on women working in management positions at large companies (MOL, thyssenkrupp, Bosch) or in the university sector. The second and third groups were selected based on personal acquaintance.

The very first focus group discussion took place on 4 March 2022, via a video call through the Microsoft Teams interface with 5 participants. They were all women aged between 35 and 50 working in senior management positions. The second focus group involved 5 participants, 3 boys, and 2 girls. Their ages ranged from 23-37 years, and all worked in SMEs. The interview took place face-to-face in the meeting room of one of the companies on 6 March 2022. In the third focus group interview, we interviewed people who work for a multinational company (Vodafone, Telekom, etc.), SMEs, or their businesses. The participants only knew that they would be asked about their attitudes towards urban transport and sustainability, but we did not mention scooters because we wanted to monitor how they reacted to this new phenomenon in the interview situation.

Thus, in this exploratory part of our research, we conducted a total of three focus groups in 2022, where stakeholders who do not use electric scooters were asked about their attitudes, experiences, and thoughts on the integration of these vehicles into transportation. The three groups were sufficiently demographically heterogeneous to represent public opinion. During the third focus group interview, the level of theoretical teasing was reached, and no substantially new information was presented in the discussion. The details of these interviews are shown in Table

1. The focus group titles shown in the first row of the table are only the authors' afterthoughts, based on the opinions they expressed.

Table 1
Focus groups details

The invented name of the focus group:	The empathic ladies	The skeptics	The 'open-minded'
The number of participants:	5 people (5 Female)	4 people (2 Male, 2 Female)	6 people (4 Female, 2 Male)
Ages:	25-45	28-37	35-45
City:	Budapest	Budapest	Wien, Budapest, Kecskemet, Szigetszentmiklós
Platform:	Ms Teams	Personal	Ms Teams
Duration:	60 min	45 min	80 min
Date:	March 4, 2022.	Marc 6, 2022.	March 8, 2022.

4.1. Research Questions

We used a pre-planned guideline with semi-structured open-ended questions for the focus group interviews. Table 2 shows the different types of questions. Starting with the broader topic of sustainability in general, we refined the questions down to their thoughts toward electric scooters and eventually asked them to give their suggestions on how they would integrate this new type of vehicle into urban transport.

Table 2
Research question types

Opening questions	Please briefly introduce yourself. What does sustainability mean to you? How do you do it in your everyday life?
Introductory questions	What mode of transport do you use to get to work? Why did you choose this? If it is not easily available to you, what other method of transport do you choose?
Transition questions	What types of shared electric vehicles are currently available in Hungary (car, scooter, motorbike, bicycle)? Have you used any of these? Are there any that you use regularly? If yes, why? What do you think about them?
Key questions	What do you think about electric scooters? What does an electric scooter mean to you? Have you ever tried them? Will you ever try them? What equipment and rules would you need to ever try an electric scooter? Please, don't hold yourself back!
Ending questions	Do you have anything else to say?

4.2. Coding

The data analysis method was conducted by hand rather than using a computer-assisted qualitative data analysis software tool, which allowed us to completely immerse and engage with the material (Fielding-Lee, 1995) and appreciate any subtle nuances it included. From the transcripts of the focus group interviews, we first highlighted the key concepts and narratives that were expressed during all three interviews and then used them to construct a nomological network of the phenomenon.

4.3. Ethic

Before attending the focus group, participants were asked to sign a consent form. Each participant was allowed to read the consent and ask questions once they arrived at the focus group. When they were ready, each participant signed a formal consent form.

5 Results of the Nomological Network

The results are presented in the logical order specified in the nomological network. However, the unique characteristics and typical statements are taken from the entire interview because additional comments and thoughts about a single topic were provided in numerous questions. The traits represent factors that are consistent across the three interviews.

5.1. Attitudes toward Electric Scooters

Using an electric scooter is difficult. The vast majority of people dislike it. On the one hand, they consider the risk of an accident, and the storing and charging are unpredictable. It is expected that a change will only happen when scootering becomes a part of daily life and pedestrians and motorists or car drivers become familiar with people who use this mode of transportation. The legislation is expected to include establishing the necessary driving license and protective equipment, as well as speed limits for scooters.

According to conversations, the lack of specific regulations on where scooters can and cannot be placed is a continuous challenge for all road users: Illegally parked scooters, sometimes littered and leaning against the sidewalk or bike path, are a common sight on the streets. Conflicts over space were thought to be one of the most relevant issues highlighted in the three interviews, both before and after the introduction of e-scooters, particularly in the framework of a race between e-scooter users and non-users (cyclists/pedestrians). The main point of contention was whether e-scooters should be prohibited from pedestrian areas and sidewalks. Even

though, the lack of dedicated infrastructure for e-scooters was highlighted by some interviewers as an issue of space distribution, also in comparison to the generous allocation of space to the car for both driving and parking.

Irresponsible riding, littering, and vandalism are important sources of conflict in many answers. 'Irresponsible riding' includes riding at crazy speeds, riding recklessly, and riding on sidewalks or pedestrian areas where it is not permitted. Cluttering refers to unplanned parking or e-scooters 'left' on pavements. Vandalism includes both intentional destruction, such as throwing e-scooters into lakes and rivers, and more common kinds of ransacking.

Based on the interviews we discovered that several problems appeared unexpectedly. This includes vandalism, e-scooter lifetime, and the requirement to recycle or dispose of e-scooters.

The focus group discussions made clear that the more people use micro-mobility devices, the more car traffic in city centers can be reduced; but like with any transportation system, this requires the necessary infrastructure. Another important consideration that comes from the interviews is that commuters should be able to travel from railway or bus stations not only by vehicle or public transportation but also by shared systems such as scooters and bicycles.

Many respondents noted a desire to investigate or use electric scooters, but only for short distances that could be covered on foot. For example, in office parks, between buildings, or from transportation hubs to office buildings. However, the following is required: appropriate infrastructure, secure, well-designed docking stations capable of storing several scooters and fitting in restricted spaces, using renewable energy sources to charge.

5.2. Attitudes toward Alternative Transportation Modes

All respondents from the first focus group ('emphatic ladies') use public transport to get to work because they do not have to sit in traffic jams, do not have to look for parking spaces, and can do other useful activities during the journey (e.g., reading, working, studying, etc.).

Three participants in the second group ('skeptics') go to work every day in an office complex in an outer district, so they usually drive. The fourth person has his own business but often must transport machinery and other goods, so he also mainly travels by car. They use car-sharing services for leisure activities (e.g., going to a restaurant in the evening). When asked, they said that they do not use public transport at all because it takes too long to get to work and is crowded and inconvenient.

Many of the participants in the third focus group ('open-minded') mentioned that they ride their bikes to work, not only because it saves time, but also because cycling

is a great activity. The only drawback is that it is not very practical in winter or rainy weather.

During the discussion, car-sharing networks were mentioned several times as a possible sustainable transport alternative. At the same time, respondents know that the operators of these sharing systems are also businesses, and they aim to make as much profit as possible. They are constantly expanding their fleet, often even rearranging cars/scooters at night. The question is how sustainable they are.

Based on the information presented above, we discovered that electric cars are also a popular mode of transportation among participants because they can be charged for free in the office building, they do not pollute the environment, they are very comfortable, and they are already covering more and more kilometers without needing to be charged.

5.3. Attitudes toward Government Policies

Generally, the respondents are not satisfied with the current rules. The reason for this, according to them, is that there is no clear definition of the means of transport (is this now a means of transport classified as a moped, a motorbike, or a bike?), no uniform definition of parking, and no uniform definition of the appropriate speed. In the case of electric scooters, they believe that it is essential to define where and what are the speed limits for using this device. In the interests of safety, it would also be possible to define the various specifications and protective equipment since: *'Typically, these are devices that can be used at speeds of up to 25 km/h, and in extreme cases at speeds of over 50 km/h.'* They believe that not only the issue of defining the right speed but also the issue of parking is not solved now. There are districts where the government has already started to identify so-called micro-mobility points, where it is obligatory to park electric scooters. Most of them argue that the painted micro-mobility points will not solve the chaos caused by e-scooters, they will just reduce it to a smaller area and then maybe the scooters will not be thrown away anywhere in the streets, but only around the micro-mobility points. This raises another problem, would anyone who wants to use one, pick from the hundreds of littered scooters? According to the participants, the problem of dropped scooters can be temporarily solved by the establishment of micro-mobility points, but this will make parking in the already crowded city center more difficult for local municipalities or make already narrow sidewalks even narrower. In addition, if it is not mandatory, users will not take their scooters there or, if they do, they will throw them on top of each other, which is not very aesthetically attractive and will damage the scooters and shorten their lifetime.

5.4. Attitudes toward Sustainability

We can assume from the responses from the three focus groups that sustainability and everyday goals toward sustainability are becoming increasingly significant in

the lives of consumers. Interviews revealed that respondents are aware that electric scooters also pollute the environment. E-scooters' carbon-dioxide emissions, like the other types of electromobility, are principally caused by manufacture, servicing, and daily transportation for maintenance and charging.

However, they think that any type of electric vehicle is better for the urban environment because it makes a difference whether we must breathe dusty, sooty, smokey air and live in continuous noise or live in cleaner air and quieter cities. In comparison to electric cars, the importance of energy supply for shared e-scooter charging is relatively modest.

They know that there are still bad practices that lead to unnecessary carbon-dioxide emissions, such as, for example, in districts of Budapest that take e-scooters off the streets every night even in cases where they are fully recharged. The participants proposed that more environmentally friendly methods of collecting the scooters for maintenance and charging are required. In general, they mentioned that using removable batteries and electrified vehicles for collection, as well as dispersed charging infrastructure, are operation characteristics that contribute to emissions reduction.

They know that the production of lithium-ion batteries for electric scooters and bicycles has a significant environmental impact through the mining of raw materials, so recycling battery cells can be an important milestone in minimizing or eliminating carbon emissions. The primary source of recycling could be from batteries in electric vehicles, as they are typically taken out of service when the vehicle's performance starts to fall short of the predicted range. At this point, the battery is still at 70-80% charge. The biggest challenge of recycling is that these batteries are made up of hundreds of lithium-ion battery cells, each of which ages in different ways, and like a single link in a chain, the weakest battery typically reduces the capacity of the whole system. Precisely for this reason, the major question for them is what will happen to use lithium-ion batteries.

6 Implications

In Budapest, there are districts from which e-scooters are banned until the issue of storage and maximizing top speed is resolved in a way that is satisfactory for all stakeholders. This is illustrated in Figure 1.

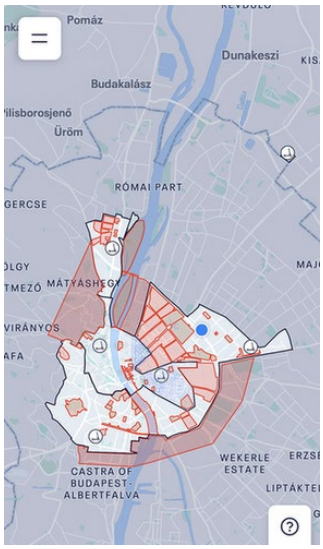


Figure 1

The districts in Budapest where you cannot stop the app and park your e-scooter

Other municipalities in Hungary, by contrast, have decided to allow scooters into their districts only if the service operators arrange charging and storage. Therefore, a network of scooter and bicycle parking, known as micro-mobility points, has been set up in the city center, first in districts V, VI, VII, VIII, and XIII of Budapest. The micro-mobility points will be designated by the local authority in the public space and will charge a fee for their use to the scooter and bicycle companies using them, at least in the frequented areas. E-scooter riders can only complete their journey and drop off their vehicles inside the parking area because the built-in GPS does not allow them to continue their journey elsewhere. Anyone who does not return it will be charged 10.000 HUF by the operator, which will be billed to the user. In many places, there will initially be virtual zones without signage, where vehicles can be parked using GPS. Using GPS, scooter operators can also set up zones where they will only allow scooters to travel at reduced speeds or switch off the electric motor.

An integrated transport service platform is also being developed, which will be a mobile phone application showing where the micro-mobility points are, how many free scooters, bikes, scooters, or public cars are available, and even how many can be booked. The app can also be used to plan a route, in principle by comparing all shared vehicles and public transport to find the optimal option. The micro-mobility points represent a major step towards the wider adoption of micro-mobility devices.

In addition, a Hungarian startup has also taken the parking and charging of electric scooters to heart. They design and manufacture electric scooter charging and docking stations that can be used to organize the devices in public areas while

reducing the charging costs and emissions for the operators. The start-up aims to build a national network of charging and docking stations in the next 2-3 years, in cooperation with local cities, municipalities, transport providers, and e-roller service providers. This would involve hundreds or even thousands of docking and charging stations, ensuring a tidy urban landscape while optimizing costs for operators.

Conclusions

While the advantages of micro-mobility and emerging flexible modes such as e-scooters are obvious, they also represent regulatory issues for municipalities. Concerns over safety, vandalism, and regulatory uncertainty have cast some doubt on the new mode's long-term feasibility. Until recently, e-scooters operated in a relative 'grey zone' in various places throughout the world since they were not included in the existing vehicle law.

In the theoretical introduction, we discussed the important factors on which the qualitative research described here was founded. The relevance of our study is not only given by the current economic situation (high petrol prices) but also by the significant social and even environmental benefits that can be derived from the decision-maker's understanding of users' expectations and attitudes towards electric scooters. The examples presented in different publications, different perspectives, and the results of our research show a wide range of attitudes of respondents towards sustainability, alternative transport, and electric scooters.

In the search for a solution, we explored the needs of our respondents regarding the infrastructure and regulation of electric scooters: we highlighted the most important measures and regulations for the target group – which, if in place, would make them willing to use electric scooters – to offer decision-makers the opportunity to integrate an ideal e-scooter ecosystem into the urban environment, contributing to a more orderly urban landscape and a positive image of electric scooters.

Our research focused on non-user perceptions of the use and limitations of shared e-scooters in Hungary. According to the interviews and our nomological network, the social construction of the meaning of electric scooter is shaped by:

- the technological affordances: such as GPS-based mobile apps and docking-charging stations in between the public transport and the free P+R stations,
- the business ecosystem: the emergence of electric scooters in urban transport has made it necessary to build new infrastructure and thus a new type of mobility solution, as the storing and the charging of scooters had to be provided,
- the government policies: the most important thing would be for the government to regulate intoxicated e-scooter riding, the age limits for using e-scooters, the number of people who can use an electric scooter at the same time, and the use of mobile phones while scootering,

- the user types/attitudes: in terms of impressions of the respondent's everyday mobility or for society and people's everyday mobility, non-users typically perceive e-scooters as harmful or wholly harmful, demonstrating significant differences in attitudes between users and non-users. The most common reason they are not using scooters is that they are satisfied with present modes of transportation and not seeing a clear necessity, as well as do not feel safe using an e-scooter in Hungary.

There are several implications for further development. First, we cannot emphasize enough how significantly the adaptation challenge relates to urban infrastructure and space allocation. As mentioned in the interviews, comprehensive policy design processes are required. These parts should be co-developed in collaboration with a variety of stakeholders, to facilitate discussion of various policy measures, recognizing their effectiveness in terms of behavioral change as well as their implementability. Based on the focus group interviews, optimal policy design would rely on national-level regulation of aspects such as drunk riding, speed, and user age, multi-stakeholder campaigns, particularly targeting non-cooperative behavior, development of education programs for all mobility system users, rules for usage/parking in specific urban areas, and further development of user recognition and verification technology in the e-scooter vehicle and associated digital platform.

On a theoretical level, several underlying questions in the governance culture have emerged, as has been typical when developing technologies deployed in a community. Higher reflection on the concept of developing technology and confrontational discussion, adaptability, and responsibility with engaging less but smartly, as well as long-term organizational learning based on experimentation and evaluation, will be required in governance culture. Simultaneously, there is an essential question about the hierarchical position of e-scooters with other modes of transportation in Hungary.

References

[1] Aminoff, A., & Sundqvist-Andberg, H. (2021) Constraints leading to system-level lock-ins—The case of electronic waste management in the circular economy. *Journal of Cleaner Production*, *322*, 129029. https://doi.org/10.1016/j.jclepro.2021.129029

[2] Benbasat & Zmud. (2003) The Identity Crisis within the Is Discipline: Defining and Communicating the Discipline's Core Properties. *MIS Quarterly*, *27*(2), 183. https://doi.org/10.2307/30036527

[3] Blumer, H. (1986) *Symbolic interactionism: Perspective and method*. University of California press

[4] Flick, U. (2014) *An introduction to qualitative research* (Edition 5). Sage

[5] Fielding, N. G., & Lee, R. M. (2002) New Patterns in the Adoption and Use of Qualitative Software. *Field Methods*, *14*(2), 197–216. https://doi.org/10.1177/1525822X02014002005

[6] Geels, F. W., Sovacool, B. K., Schwanen, T., & Sorrell, S. (2017) Sociotechnical transitions for deep decarbonization. *Science*, *357*(6357), 1242–1244. https://doi.org/10.1126/science.aao3760

[7] Gössling, S. (2020) Integrating e-scooters in urban transportation: Problems, policies, and the prospect of system change. *Transportation Research Part D: Transport and Environment*, *79*, 102230. https://doi.org/10.1016/j.trd.2020.102230

[8] Gössling, S., & Cohen, S. (2014) Why sustainable transport policies will fail: EU climate policy in the light of transport taboos. *Journal of Transport Geography*, *39*, 197–207. https://doi.org/10.1016/j.jtrangeo.2014.07.010

[9] Hardt, C., & Bogenberger, K. (2019) Usage of e-Scooters in Urban Environments. *Transportation Research Procedia*, *37*, 155–162. https://doi.org/10.1016/j.trpro.2018.12.178

[10] Kopplin, C. S., Brand, B. M., & Reichenberger, Y. (2021) Consumer acceptance of shared e-scooters for urban and short-distance mobility. *Transportation Research Part D: Transport and Environment*, *91*, 102680. https://doi.org/10.1016/j.trd.2020.102680

[11] Latinopoulos, C., Patrier, A., & Sivakumar, A. (2021) Planning for e-scooter use in metropolitan cities: A case study for Paris. *Transportation Research Part D: Transport and Environment*, *100*, 103037. https://doi.org/10.1016/j.trd.2021.103037

[12] Liamputtong, P. (2011) *Focus Group Methodology: Principles and Practice*. SAGE Publications Ltd. https://doi.org/10.4135/9781473957657

[13] Ma, Q., Yang, H., Ma, Y., Yang, D., Hu, X., & Xie, K. (2021) Examining municipal guidelines for users of shared E-Scooters in the United States. *Transportation Research Part D: Transport and Environment*, *92*, 102710. https://doi.org/10.1016/j.trd.2021.102710

[14] Manskie, J., 2019. 'The motorized scooter boom that hit a century before dockless scooters' Available at: https://www.smithsonianmag.com/history/motorized-scooter-boom-hit-century-dockless-scooters-180971989/

[15] Schellong, D., Sadek, P., Schaetzberger, C., Barrack, T., 2019. The Promise and Pitfalls of E-Scooter Sharing. Boston Consulting Group. Available online: https://www.bcg.com/publications/2019/promise-pitfalls-e-scooter-sharing

[16] Shaheen, S. (2019) *Micromobility Policy Toolkit: Docked and Dockless Bike and Scooter Sharing*. https://doi.org/10.7922/G2TH8JW7

[17] Semenov, A., 2017. Why Will Micro Mobility Industry Make the Future? Available online: https://medium.com/@Splyt/why-will-micro-mobility-industry-make-the-future-1b0a628ae3d0

[18] Weschke, J., Oostendorp, R., & Hardinghaus, M. (2022) Mode shift, motivational reasons, and impact on emissions of shared e-scooter usage. *Transportation Research Part D: Transport and Environment*, *112*, 103468. https://doi.org/10.1016/j.trd.2022.103468

[18] Wilkinson, S. (1999) Focus Groups: A Feminist Method. *Psychology of Women Quarterly*, *23*(2), 221–244, https://doi.org/10.1111/j.1471-6402.1999.tb00355.x

Thermal Property Investigation of a Simplified, Dye-Sensitized Solar Cell: Analyzing the Heat Transfer in Solids and Fluids, with a Turbulent Flow, for Open-Circuit Voltage Calculations

Ervin Racz

Kando Kalman Faculty of Electrical Engineering, Obuda University, 1034 Budapest, Hungary, racz@uni-obuda.hu

Abstract: Utilizing renewable energy sources is a long-term mitigation solution. Electricity generation by photovoltaic (PV) is growing rapidly and PV is at the forefront of the renewable energy revolution. Dye-Sensitized Solar Cell (DSSC) is one of the more popular transparent solar cells used for architectures. Additionally, DSSC is a cost-effective solar cell and simple to manufacture because of the inexpensive materials, which are used in the fabrication processes. This article explores a method to optimize the thermal properties of the DSSC. For the simulation COMSOL Multiphysics software usage has been used and 3-dimensional scheme via numerically solved coupled models – such as heat transfer in solids and fluids, and turbulent flow – in stationary mode have been applied. Boundary conditions have been added and described in great details. For the optimization process, the glass substrate has been modified. The temperature distribution in the volume and in the surface; pressure distribution in the volume; and velocity distribution in the volume have been presented. On the other hand, the temperature distribution, along a line, is extracted from the 3-dimensional geometry, to analyze the thermal behavior and use for further investigation, the electrical behavior of the Dye-Sensitized Solar Cell. The results show that the higher temperature, which is located at the top part of the geometry, is 318 K. The calculated open-circuit voltage of the cell is between 448 mV and 452 mV and the gradient is $-2\ mV/°C$.

Keywords: Dye-Sensitized Solar Cell; Optimization; Thermal Analysis; Simulation; Open-Circuit Voltage

1 Introduction

The Global Energy Crisis is a high-level scientific challenge, which has not been solved. It is an intensely debated topic, concerning what kind of measurements have to be taken, in order to reduce carbon-dioxide emission into the atmosphere [1]. Devastating results are forecasted, if the current trends continue in electricity

production [2] [3]. Many researchers emphasized that the use of cost-effective, low-carbon energy technologies are very relevant and undisputedly important [4]. Furthermore, if this continues, carbon-dioxide emissions could be 60% higher within 25 years compared to the current emissions. The most significant greenhouse gas is carbon-dioxide. On the other hand, the energy demand is predicted to be 70% higher by 2050 [5]. Fossil fuel-based energy generation has been the dominant mode of energy production in previous decades, contributing a great deal of greenhouse gas emissions [6-8].

Utilizing renewable energy sources is a long-term mitigation solution. As a matter of fact, the electromagnetic radiation (waves) hit the earth's surface can significantly cover the needs of humans' activity [9]. This is why, solar energy is more urgent than ever before. On the other hand, billions of US dollars had been invested into clean energy development [10]. Moreover, electricity generation by photovoltaics (PV) is growing rapidly and PV is at the forefront of the renewable energy revolution [11]. The device harnesses the power of the electromagnetic wave and directly converts it into electrical energy. However, several solar cell types exist, the most commonly used photovoltaic device is the monocrystalline solar panel. The principle of the effect is that the striking photons excite the electrons in the semiconductor material, causing the generation and the flow of the electric current, known as direct current (DC) [12]. After that DC is converted into alternating current (AC) using inverters [13]. The current trends show that solar cell research also moved towards building integrated photovoltaic in which transparent solar cell is favorable [14].

Dye-sensitized solar cell (DSSC) is one of the most popular transparent solar cells used for architecture. Additionally, DSSC is a cost-effective solar cell and simple to manufacture solar cell because inexpensive materials are used for fabrication processes. DSSC is composed of five main components: glass substrate coated with transparent conductive oxide, titanium-dioxide semiconductor, dye sensitizer, electrolyte solution and platinum catalyst. The sensitizing dye molecules are excited by the incoming photons, then electrons injected from dye molecule into the conduction band of the titanium-dioxide semiconductor layer. The charge separation and charge generation are done separately. Moreover, after the injection, the dye is oxidized and it can receive electrons from the electrolyte solution. After, ions diffuse to the counter electrode in order to receive electrons [15]. The dye absorber has to meet with the following requirement: It has to be able to absorb a wide part of the solar spectrum. Interestingly, the number of publications has increased due to the intensive research interest towards this innovative cell [16].

Arifin et al. investigated the effect of heat sink properties on solar cell cooling system using numerical simulations and experimental investigation. In the study a passive cooling system, namely heat sink, has been added. The cooling system was modelled using Navier-Stokes equation. Applying 15 pins copper-based heat sink, the temperature of the solar cell decreased by 10.2 °C, and the efficiency of the PV increased by 2.74% [17]. Other researchers conducted analysis on dye-sensitized

solar cell and temperature improvement. Baiju et al. reported a temperature sintering process of titanium-dioxide on polymer substrates using heat sink [18]. Others proposed a dye-sensitized solar cell – thermoelectric hybrid system in order to improve the output efficiency and electric power of the system [19]. On the other hand, this innovative method also enhances the thermal effect of the solar cell by recycling the accumulated heat inside the DSSC. In the study of Chang. et al, a solar thermoelectric module investigation was doing by the integration of CuO thin films [20]. Chen et al. analyzed the thermal performance of a DSSC module using numerical model. Their results indicate that the wind speed affect the cooling effect: the greater the wind speed is, the better the cooling effect [21].

Furthermore, thermal stress analysis also takes an import role in the investigation of the dye-sensitized solar cell. Saasaoui et al. studied the electrical characterization of natural DSSC applying different thermal stress conditions [22]. Investigation under extended thermal stress was also conducted by Yadav et al. [23]. Bari et al. also investigated the thermal stresses of the DSSC device by using climatic chamber at a constant temperature. Their investigation showed that the degradation reduces the photogenerated current [24]. Solanki et al. investigated a concentrator PV module, in which the cell is integrated in V-troughs, resulting better heat dissipation [25]. Kim et al. improved heat dissipation in crystalline silicon PV by using a highly thermal conducting backsheet [26]. Shao et al. investigated the optical and thermal performance of dynamic concentrating solar module [27].

In previous publications, dye-sensitized solar cell – thermoelectric generator (DSSC-TEG) hybrid system, was investigated and the results showed that the use of TEG dissipated heat from the backside of the solar cell by temperature difference in the two sides. This way the thermoelectric generator can increase the efficiency of the system [28]. This article explores another method optimizing the thermal properties of the dye-sensitized solar cell. The aim of this article is to simulate the temperature distribution of the glass substrate of a dye-sensitized solar cell by modifying the glass structure with heat dissipation spheres. Applying this feature, DSSC remains transparent, and its heat dissipation can be increased. This study paves the way in providing some key insights related to this field.

2 Modelling and Simulation

During the simulation, the thermal analysis of the dye-sensitized solar cell was investigated by modifying the glass substrate. As a first step, the geometry and material were selected, then adding the physics and the mesh. As a next step, the simulations were performed, data were collected, results were analyzed, and conclusions were drawn.

COMSOL Multiphysics software was used to build the model and simulate it. COMSOL is often used for thermal analysis of various devices [29]. The simulation was in three-dimensional scheme via numerically solved coupled models. On the other hand, in the simulation steady state conditions have been applied which means that for the heat transfer the temperature does not change with time; for the fluid dynamics, the fluids flow does not change with time. In many studied literature, stationary condition is favorable for the simulations [17] [29].

2.1 Geometry and Materials

As it was mentioned, the conventional dye-sensitized solar cell contains five main components. In the preparation of the simulation design, some simplifications were made: DSSC solar cell was substituted by the glass carrier because the glass carrier constitute a significant portion of the mass of the DSSC.

During the definition of geometries, it was taken into account that the research could be continued further investigation in the future (e.g., comparison with another geometry structure). Therefore, the dimensions of the glass block (represents the dye-sensitized solar cell) are 6 mm × 4 mm × 4 mm. The dimensions are given in the following format: Width × Depth × Height. The glass block is in the middle, and it is surrounded by a block of air. At the bottom of the glass carrier, heat dissipation balls are located, with diameters of 1 mm. Thus, the system remains transparent. The glass block and the six pieces of spheres are built together. In other words, the number of the dissipation balls are six.

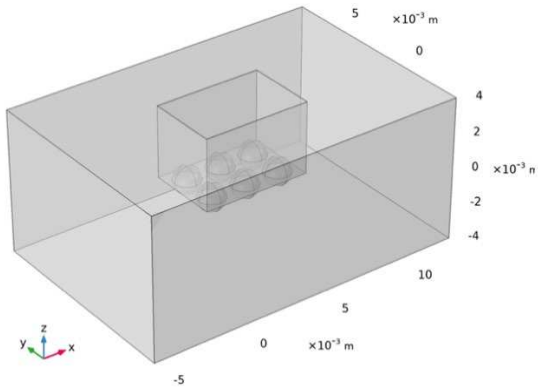


Figure 1

The designed and built geometry in three dimensions. The glass block is in the middle, and it is surrounded by the air. Also, at the bottom of the glass, heat dissipation balls are located.

The following aspects have been taken into consideration in the selection of dissipation ball diameter:

(i) The 1 mm diameter can provide sufficient surface area for efficient heat dissipation without affecting the materials' transparency

(ii) The computational time of the simulation is shorter with 1 mm diameter, with larger balls, the computational load would increase

(iii) More dissipation balls can be placed under the surface. The dimension of the air block is 18 mm × 12 mm × 8 mm. Figure 1 represents the designs and the built geometry in three dimensions, where the dimension of the three axes, are set in mm.

The air is taken into account as an incompressible fluid if the flow velocity is below 100 m/s [17] [30]. The selected materials, such as glass and air, are from COMSOL library. These built-in-materials contain the dynamic viscosity, ratio of specific heat, heat capacity at constant pressure, density and thermal conductivity.

2.2 Heat Transfer

The heat transfer module has been added to physics. From the heat transfer module, heat transfer in solids and fluid joint interfaces have been chosen.

2.2.1 Heat Transfer in Solids

The heat conduction in solid is described by the Fourier equation, where the heat transfers from the higher temperature point to the lower temperature point. The accumulated heat inside the system has to be transported in order to reduce the temperature of the cell. Thus, simulating the heat transfer in solids are essential. The following equation describes the heat transfer in solid:

$$\rho C_p \left(\frac{\partial T}{\partial t} + u_{trans} \cdot \nabla T\right) + \nabla \cdot (q + q_r) = -\alpha T : \frac{dS}{dt} + Q \quad (1)$$

Where,

- ρ is the density
- C_p is the specific heat at constant stress
- T is the absolute temperature
- u_{trans} is the velocity vector of translation motion
- q is the heat flux by conduction
- q_r is the heat flux by radiation
- α is the coefficient of thermal expansion
- S is the second Piola-Kirchhoff stress tensor
- Q contains additional heat sources

From equation (1), it is seen that the temperature and the second Piloa-Kirchhoff stress tensor are time dependent. Taking into account that the simulation is stationary, these elements in the equation is considered to zero. In other words, the first derivative of the temperature with respect to time is zero, and the first derivative of the second Piloa-Kirchhoff stress tensor with respect to time is zero. Surface-to-Ambient Radiation boundary condition has been applied between the solid surface and the fluid surface contact. The surface emissivity was 0.85 [31]. Moreover, heat flux boundary condition has been added and general inward heat flux was selected and the value of it is 1000 W/m^2.

2.2.2 Heat Transfer in Fluids

Taking into consideration the solar cell is surrounded by air, the simulation of heat transfer in fluids are also crucial. Compressible and incompressible fluids are distinguished. If the flow velocity of the fluid is below 100 m/s, the air can be considered as incompressible fluid [17], which means that the density of the fluid is constant. In other words, the density of the fluid does not change in time and with the flow. On the other hand, the following equation, equation (2), describes the incompressible fluid in three-dimensional version:

$$\frac{\partial \rho}{\partial x} = 0; \frac{\partial \rho}{\partial y} = 0; \frac{\partial \rho}{\partial z} = 0 \tag{2}$$

Where,

- ρ is the density
- x, y, z are the three-dimension coordinates

$$\rho C_p \left(\frac{\partial T}{\partial t} + u \cdot \nabla T\right) + \nabla \cdot (q + q_r) = Q_p + Q_{vd} + Q \tag{3}$$

where $Q_p = \alpha_p T \left(\frac{\partial p}{\partial t} + u \cdot \nabla p\right)$ and $Q_{vd} = t{:}\nabla u$, where,

- The dependent variables are the temperature, T
- Pressure, p

In equation (3), ρ represents the density:

- C_p is the specific heat capacity at constant pressure
- u is the velocity vector
- α_p is the coefficient of thermal expansion
- q is the heat flux by conduction
- q_r is the heat flux by radiation
- τ is the viscous stress tensor
- Q is the heat source

On the other hand, $\alpha_p = -\frac{1}{\rho}\frac{\partial \rho}{\partial T}$. Furthermore, the heat flux boundary condition has been added to the simulation. Convective heat flux has been chosen and the heat transfer coefficient was set to be 5 W/(m^2 K) and the external temperature was set to ambient temperature. The time-dependent variables appealing in the third equation are also zero, because the simulation is solved in stationary mode.

2.3 Turbulent Flow

From the single-phase flow interfaces, turbulent flow, $k-\varepsilon$ interface was selected. The program solves the Raynolds-averaged Navier-Stokes equations for the velocity field, pressure and the turbulent viscosity. According to [32], the $k-\varepsilon$ model is one of the most used turbulent models for industrial applications. The model contains two dependent variables which are the turbulent kinetic energy, namely k, and the turbulent dissipation rate, known as ε; and two additional transport equation, equation (5) and equation (7).

The turbulent viscosity is modelled as:

$$\mu_T = \rho C_\mu \frac{k^2}{\varepsilon} \tag{4}$$

where C_μ is the model constant.

On the other hand, the transport equation for k is the following:

$$\rho \frac{\partial k}{\partial t} + \rho u \cdot \nabla k = \nabla \cdot \left(\left(\mu + \frac{\mu_T}{\sigma_k} \right) \nabla k \right) + P_k - \rho \varepsilon \tag{5}$$

where the production term i:

$$P_k = \mu_T \left(\nabla u : (\nabla u + (\nabla u)^T) - \frac{2}{3}(\nabla \cdot u)^2 \right) - \frac{2}{3}\rho k \nabla \cdot u \tag{6}$$

The transport equation for ε reads:

$$\rho \frac{\partial \varepsilon}{\partial t} + \rho u \cdot \nabla \varepsilon = \nabla \cdot \left(\left(\mu + \frac{\mu_T}{\sigma_\varepsilon} \right) \nabla \varepsilon \right) + C_{\varepsilon 1} \frac{\varepsilon}{k} P_k - C_{\varepsilon 2} \rho \frac{\varepsilon^2}{k} \tag{7}$$

Constant values are determined in the moles such as $C_\mu = 0.09$, $C_{\varepsilon 1} = 1.44$, $C_{\varepsilon 2} = 1.92$, $\sigma_k = 1.0$, $\sigma_\varepsilon = 1.3$

Furthermore, two boundary conditions have been applied: (i) inlet boundary condition and (ii) outlet boundary condition. Similar to the previous points, the current equation was also solved in a stationary mode.

2.4 Mesh

After defining the parameters, building the geometries, selecting the used materials and applying the useful physics, the next step was to set the mesh. In the mesh builder, physics-controlled mesh was selected under sequence type and the element size was set to normal.

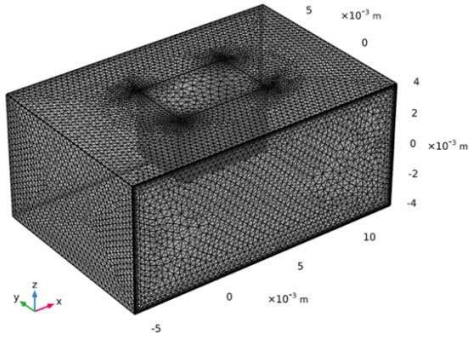


Figure 2
Mesh about the built geometry. The element size was set to normal and it is seen that the element size is much dense at the connections.

Figure 2 represents the mesh about the built geometries. The choice of the element size was made because of the computation time and computation capacity had to be taken into account. Moreover, the mesh is dense at the connection points. For the simulation Dell laptop with Intel64 Family 6 Model 141 Stepping 1, 6 cores CPU with 8 GB RAM was used. The computation time was 1 hour 59 minutes and 40 seconds.

3 Results and Discissions

The simulation aimed to explore and analyze the temperature distribution of the glass substrate of a dye-sensitized solar cell by modifying the glass substrate with heat dissipation spheres. In the following subsections, results of the temperature distribution will be shown, then the data preprocessing comes. Additionally, results from the simulation will be used for further dye-sensitized solar cell simulation. The findings provide a comprehensive understanding of the system's behavior and pave the way for potential further research in the dye-sensitized solar cell optimization filed.

Simulations were run without heat dissipation spheres at the back contact. Comparing the results, it can be stated that the dissipation balls assist in heat dissipation since the volume temperature becomes lower (from $\sim 319\ K\ to \sim 317\ K$) in stationary mode.

3.1 Results of the Simulation

The temperature distribution in the volume and in the surface, pressure distribution in the volume and velocity distribution in the volume will be presented in this subchapter. Figure 3 shows the temperature distribution across the three-dimensional geometry with the streamline.

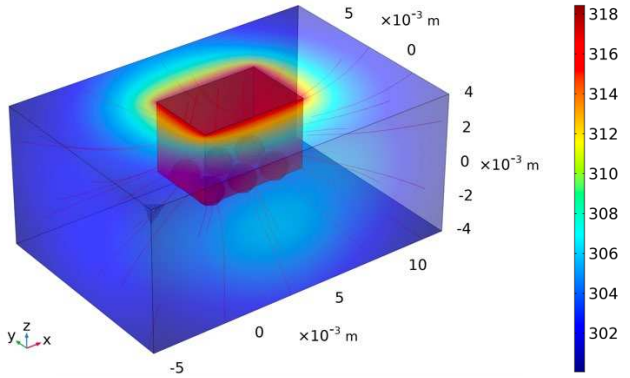


Figure 3
Temperature distribution across the three-dimensional geometry with streamline. Streamlines represent the flow of the heat.

It is visible from the figure that the central part heats the entire geometry and the heat transports from the higher temperature point to the lower temperature point. Also, in the figure, the streamlines are visible and show the direction of the heat. On the right side of the figure, the color legend, and on the left side of the figure the axis orientation can be seen. The dimension of the temperature is in kelvin. Furthermore, because of the glass substrate surrounded by the air, the air can dissipate a great part of heat from the glass substrate.

On the other hand, the glass body is higher than 316 K. Moreover, the 2D surface plot represents a cross-section view of the geometry. The mentioned surface plot is shown in Figure 4, and it shows the temperature distribution in two dimension.

It is worth mentioning that the surface plot represents the temperature distribution of the geometries better. Additionally, the added spheres contribute to the greater heat dissipation. The color mapping in both Figures 3 and 4 plots help visualize the intensity of the temperature variable indicating the warmer temperature points and the colder temperature points. For the visualization XZ-planes have been added.

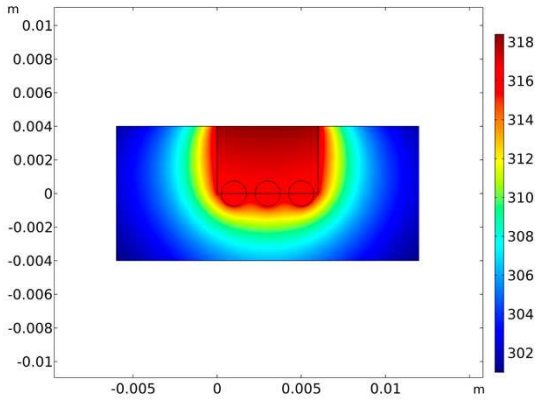


Figure 4

A cross-section wire of the geometries indicating the temperature values via the color bar on the right side of the figure. The figure represents the XZ-cut plane section of Figure 3 plot.

Moreover, in order to enhance the evaluation of the results, the pressure distribution plot can be observed in Figure 5. Additionally, it shows how the pressure changes within the system.

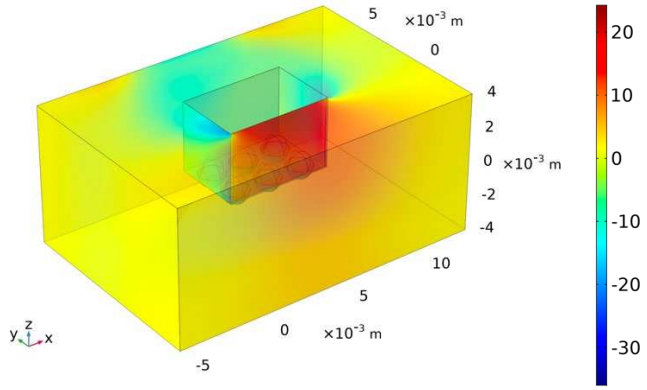


Figure 5

Pressure distribution across the three-dimensional geometry. On the right side, the color bar is visible. (Reddish color means warmer areas.)

The incoming air flows in from the front side. Therefore, the pressure values are higher in the front side, representing reddish color. Moreover, from Figure 5 it can be seen that at the edges of the air block (high block surrounds the glass block), the value of the pressure is either zero or approached to zero. The other figure which represents the turbulent flow in the materials can be observed in the Figure 5.

Figure 6 illustrates the change in velocity in the geometry. The y-component of the velocity is shows in the figure. It displays how the velocity fluctuates over the material structure. Furthermore, value of the system seems to be higher at certain regions, while in other areas, the value of the velocity is lower. In those regions where the velocity is higher, those areas indicating faster motion. The visual representation of the y-component of the velocity across the geometry provides insight into the distribution of the velocity and helps to understand how the air flows in such geometries behaves throughout the designed and built system.

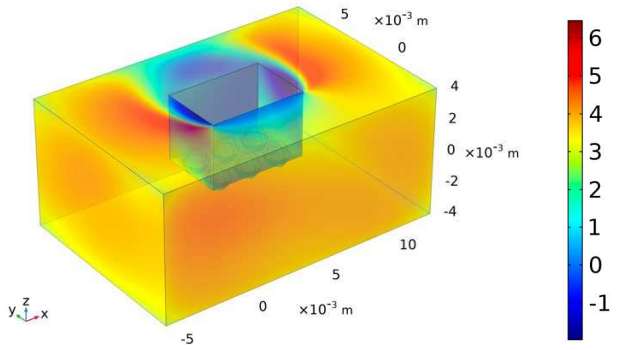
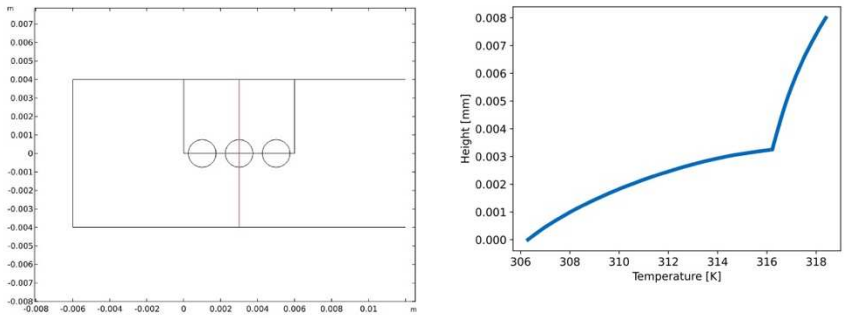


Figure 6
They y-component of the velocity across the geometry where the turbulent flow was applied. (Red color means warmer areas.)

3.2 Data Preprocessing

After the simulation and temperature distribution, pressure distribution and y-component of the velocity distribution visualizations, a cut-line 2D from the cut plane data set has been added. The starting point of the x component (point 1) of the line is 3 mm, and the finishing point of it (point 2) is 3 mm. Additionally, the starting point of the y component (point 1) is -4 mm and the finishing point of it (point 2) is 4 mm. Figure 7a, shows the aforementioned cut line used for further data visualization and Figure 7b, represents the temperature distribution across the cut line. The cut line was extracted from the three-dimensional geometry shown in Figure 7a). Moreover, Figure 7b, provides insight into how the temperature variable with distance changes. It can be seen that the part, at 0.008 mm height, has the highest temperature. The temperature gradient can be identified and also one of the key regions of interest in the plot is at 0.003, where the two geometries (glass substrate and air) meet.



a) *b)*

Figure 7

a) The cut line used for further data visualization and data analyzation, where the horizontal coordinates are the x coordinate and the vertical coordinates are the z coordinates. b) Temperature distribution across the cut plane. The horizontal coordinates are the temperature in kelvin dimension and the horizontal coordinates are the height in millimeter dimension.

3.3 Open-Circuit Voltage

Further investigations were conducted based on the temperature results obtained from the simulation. Temperature is a crucial factor of the solar cell investigation. In the [33] publication, an application has been developed to investigate the open-circuit voltage value, which takes into consideration the experimental data and the optical equations of the dye-sensitized solar cell using numerical finite element method as numerical procedure. The input parameters are the thickness of the cell, reflection, electron density in the dark, energy between the conduction band and the Fermi level, electron lifetime and the cell temperature. Furthermore, photoncount-wavelength spectrum, transmission-wavelength spectrum and the irradiation intensity-wavelength spectrum can be imported from a .txt file. The open-circuit voltage simulation was conducted based on the result from Figure 7b. In other words, 316 K and 318 K temperature values have been taken into account for further investigation. After adding the temperature value to the application, the solved open-circuit values are 448 mV if the temperature is 318 K, and 452 mV if the temperature value is 316 K. From the results it is seen that, if the temperature increases, the open-circuit voltage would decrease. Moreover, the temperature gradient is $-2\ mV/°C$.

Conclusions

This article explores a novel method to optimize the thermal properties of dye-sensitized solar cells. In an earlier publication, the optimization of the dye-sensitized solar cell was focused on applying machine learning techniques. The aim of this article is simulating the temperature distribution of the glass substrate, of a

dye-sensitized solar cell, by modifying the glass structure with heat dissipation spheres. Applying this feature, the dye-sensitized solar cell remains transparent and its heat dissipation can be increased. Also, simulations were run without heat dissipation spheres at the black contact. Comparing the results, it can be stated that the dissipation balls assist in heat dissipation (from $\sim 319\ K$ to $\sim 317\ K$), since the volume temperature becomes lower. For the simulation, the COMSOL Multiphysics Software has been used with a three-dimensional scheme, via numerically solved coupled models – heat transfer in solids and fluids, and turbulent flow – in stationary mode have been applied.

The temperature distribution in the volume and at the surface, pressure distribution in the volume and y-component of velocity distribution in the volume, have been presented. Furthermore, the temperature distribution along a line, is extracted from the three-dimensional geometry, to analyze the thermal behaviors and used for open-circuit voltage simulation. The results show that the glass body temperature is between $316\ K - 318\ K$, resulting in $448\ mV - 452\ mV$ open-circuit voltage. On the other hand, if the temperature increases, the open-circuit voltage parameter decreases. The results obtained from the simulations, provide an opportunity for enhancing the short-circuit voltage of the dye-sensitized solar cell, through the applied application.

References

[1] S. J. Davis and K. Caldeira, 'Consumption-based accounting of CO2 emissions', *Proceedings of the National Academy of Sciences*, Vol. 107, No. 12, pp. 5687-5692, Mar. 2010, doi: 10.1073/pnas.0906974107

[2] A. M. Omer, 'Energy, environment and sustainable development', *Renewable and Sustainable Energy Reviews*, Vol. 12, No. 9, pp. 2265-2300, Dec. 2008, doi: 10.1016/j.rser.2007.05.001

[3] B. Beszédes, K. Széll, and G. Györök, 'Redundant Photo-Voltaic Power Cell in a Highly Reliable System', *Electronics*, Vol. 10, No. 11, p. 1253, May 2021, doi: 10.3390/electronics10111253

[4] O. Shvets, M. Seebauer, A. Naizabayeva, and A. Toleugazin, 'Monitoring and Control of Energy Consumption Systems, using Neural Networks', *ACTA POLYTECH HUNG*, Vol. 20, No. 2, pp. 125-144, 2023, doi: 10.12700/APH.20.2.2023.2.7

[5] M.-E. Ragoussi and T. Torres, 'New generation solar cells: concepts, trends and perspectives', *Chem. Commun.*, Vol. 51, No. 19, pp. 3957-3972, 2015, doi: 10.1039/C4CC09888A

[6] A. K. Pandey, V. V. Tyagi, J. A. Selvaraj, N. A. Rahim, and S. K. Tyagi, 'Recent advances in solar photovoltaic systems for emerging trends and advanced applications', *Renewable and Sustainable Energy Reviews*, Vol. 53, pp. 859-884, Jan. 2016, doi: 10.1016/j.rser.2015.09.043

[7] B. Beszédes, K. Széll, and G. Györök, 'A Highly Reliable, Modular, Redundant and Self-Monitoring PSU Architecture', *ACTA POLYTECH HUNG*, Vol. 17, No. 7, pp. 233-249, 2020, doi: 10.12700/APH.17.7.2020.7.13

[8] I. Diahovchenko, M. Kolcun, Z. Čonka, V. Savkiv, and R. Mykhailyshyn, 'Progress and Challenges in Smart Grids: Distributed Generation, Smart Metering, Energy Storage and Smart Loads', *Iran J Sci Technol Trans Electr Eng*, Vol. 44, No. 4, pp. 1319-1333, Dec. 2020, doi: 10.1007/s40998-020-00322-8

[9] X. Li, H. Liao, Y.-F. Du, C. Wang, J.-W. Wang, and Y. Liu, 'Carbon dioxide emissions from the electricity sector in major countries: a decomposition analysis', *Environ Sci Pollut Res*, Vol. 25, No. 7, pp. 6814-6825, Mar. 2018, doi: 10.1007/s11356-017-1013-z

[10] J. Gong, K. Sumathy, Q. Qiao, and Z. Zhou, 'Review on dye-sensitized solar cells (DSSCs): Advanced techniques and research trends', *Renewable and Sustainable Energy Reviews*, Vol. 68, pp. 234-246, Feb. 2017, doi: 10.1016/j.rser.2016.09.097

[11] R. Ghabour and P. Korzenszky, 'Dynamic Modelling and Experimental Analysis of Tankless Solar Heat Process System for Preheating Water in the Food Industry', *ACTA POLYTECH HUNG*, Vol. 20, No. 4, pp. 65-83, 2023, doi: 10.12700/APH.20.4.2023.4.4

[12] N. Ngoc Son and L. The Vinh, 'Parameter Estimation of Photovoltaic Model, Using Balancing Composite Motion Optimization', *ACTA POLYTECH HUNG*, Vol. 19, No. 11, pp. 27-46, 2022, doi: 10.12700/APH.19.11.2022.11.2

[13] K. Sharma, V. Sharma, and S. S. Sharma, 'Dye-Sensitized Solar Cells: Fundamentals and Current Status', *Nanoscale Res Lett*, Vol. 13, No. 1, p. 381, Dec. 2018, doi: 10.1186/s11671-018-2760-6

[14] Y. Li, L. Li, W. Deng, D. Zhu, and L. Hong, 'Building Integrated Photovoltaic (BIPV) Development Knowledge Map: A Review of Visual Analysis Using CiteSpace', *Buildings*, Vol. 13, No. 2, p. 389, Jan. 2023, doi: 10.3390/buildings13020389

[15] M. Onodera *et al.*, 'Modeling of Dye-Sensitized Solar Cells Based on TiO_2 Electrode Structure Model', *Jpn. J. Appl. Phys.*, Vol. 49, No. 4, p. 04DP10, Apr. 2010, doi: 10.1143/JJAP.49.04DP10

[16] S. Bose, V. Soni, and K. R. Genwa, 'Recent Advances and Future Prospects for Dye Sensitized Solar Cells: A Review', Vol. 5, No. 4, p. 9, 2015

[17] Z. Arifin, S. Suyitno, D. D. D. P. Tjahjana, W. E. Juwana, M. R. A. Putra, and A. R. Prabowo, 'The Effect of Heat Sink Properties on Solar Cell Cooling Systems', *Applied Sciences*, Vol. 10, No. 21, p. 7919, Nov. 2020, doi: 10.3390/app10217919

[18] K. G. Baiju, B. Murali, R. Subba Rao, K. Jayanarayanan, and D. Kumaresan, 'Heat sink assisted elevated temperature sintering process of TiO2 on polymer substrates for producing high performance flexible dye-sensitized solar cells', *Chemical Engineering and Processing - Process Intensification*, Vol. 149, p. 107817, Mar. 2020, doi: 10.1016/j.cep.2020.107817

[19] Y. J. Kim *et al.*, 'Dye-Sensitized Solar Cell–Thermoelectric Hybrid Generator Utilizing Bipolar Conduction in a Unified Element', *ACS Appl. Energy Mater.*, Vol. 3, No. 5, pp. 4155-4161, May 2020, doi: 10.1021/acsaem.0c00513

[20] H. Chang, M.-J. Kao, K.-C. Cho, S.-L. Chen, K.-H. Chu, and C.-C. Chen, 'Integration of CuO thin films and dye-sensitized solar cells for thermoelectric generators', Current Applied Physics, Vol. 11, No. 4, pp. S19-S22, Jul. 2011, doi: 10.1016/j.cap.2010.12.039

[21] S. Chen *et al.*, 'Numerical model analysis of thermal performance for a dye-sensitized solar cell module', *J. Phys. D: Appl. Phys.*, Vol. 46, No. 48, p. 485106, Dec. 2013, doi: 10.1088/0022-3727/46/48/485106

[22] S. Saadaoui, A. Torchani, M. A. B. Youssef, and R. Gharbi, 'Fabrication, modeling and electrical characterization of natural dye sensitized solar cell under different thermal stress conditions', *Optik*, Vol. 127, No. 20, pp. 10058-10067, Oct. 2016, doi: 10.1016/j.ijleo.2016.07.092

[23] S. K. Yadav, S. Ravishankar, S. Pescetelli, A. Agresti, F. Fabregat-Santiago, and A. Di Carlo, 'Stability of dye-sensitized solar cells under extended thermal stress', *Phys. Chem. Chem. Phys.*, Vol. 19, No. 33, pp. 22546-22554, 2017, doi: 10.1039/C7CP04598K

[24] D. Bari *et al.*, 'Thermal stress effects on Dye-Sensitized Solar Cells (DSSCs)', *Microelectronics Reliability*, Vol. 51, No. 9-11, pp. 1762-1766, Sep. 2011, doi: 10.1016/j.microrel.2011.07.061

[25] C. Solanki, C. Sangani, D. Gunashekar, and G. Antony, 'Enhanced heat dissipation of V-trough PV modules for better performance', *Solar Energy Materials and Solar Cells*, Vol. 92, No. 12, pp. 1634-1638, Dec. 2008, doi: 10.1016/j.solmat.2008.07.022

[26] N. Kim, D. Kim, H. Kang, and Y.-G. Park, 'Improved heat dissipation in a crystalline silicon PV module for better performance by using a highly thermal conducting backsheet', *Energy*, Vol. 113, pp. 515-520, Oct. 2016, doi: 10.1016/j.energy.2016.07.046

[27] Z. Shao, B. Wang, L. Cai, C. Chang, and L. Sun, 'Optical and thermal performance of dynamic concentrating solar module for building integration', *Journal of Cleaner Production*, Vol. 367, p. 132931, Sep. 2022, doi: 10.1016/j.jclepro.2022.132931

[28] Z. Varga and E. Racz, 'Machine Learning Analysis on the Performance of Dye-Sensitized Solar Cell—Thermoelectric Generator Hybrid System', *Energies*, Vol. 15, No. 19, p. 7222, Oct. 2022, doi: 10.3390/en15197222

[29] P. Saxena and N. E. Gorji, 'COMSOL Simulation of Heat Distribution in Perovskite Solar Cells: Coupled Optical–Electrical–Thermal 3-D Analysis', *IEEE J. Photovoltaics*, Vol. 9, No. 6, pp. 1693-1698, Nov. 2019, doi: 10.1109/JPHOTOV.2019.2940886

[30] J.-H. Kim, S.-H. Park, and J.-T. Kim, 'Experimental Performance of a Photovoltaic-thermal Air Collector', *Energy Procedia*, Vol. 48, pp. 888-894, 2014, doi: 10.1016/j.egypro.2014.02.102

[31] R. Gardon, 'The Emissivity of Transparent Materials', *J American Ceramic Society*, Vol. 39, No. 8, pp. 278-285, Aug. 1956, doi: 10.1111/j.1151-2916.1956.tb15833.x

[32] V. Guimet and D. Laurence, 'A LINEARISED TURBULENT PRODUCTION IN THE k-ε MODEL FOR ENGINEERING APPLICATIONS', in *Engineering Turbulence Modelling and Experiments 5*, Elsevier, 2002, pp. 157-166, doi: 10.1016/B978-008044114-6/50014-4

[33] Z. Varga and E. Racz, 'Application Development with Finite Element Method to Calculate Photogeneration Rate and Open-Circuit Voltage of Dye Sensitized Solar Cell', *Syst. Theor. Control Comput. J.*, Vol. 2, No. 2, pp. 17-24, Dec. 2022, doi: 10.52846/stccj.2022.2.2.39

Extending the Input and Transformation Space of Different TP Models: An LMI-based Feasibility Analysis

Andrea Wéber

Széchenyi István University, Doctoral School of Multidisciplinary Engineering Sciences (MMTDI), Egyetem tér 1, H-9026 Győr, Hungary, Email: weber.andrea@sze.hu

Miklós Kuczmann

Széchenyi István University, Department of Power Electronics and Electric Drives, Egyetem tér 1, H-9026 Győr, Hungary, Email: kuczmann@sze.hu

Péter Baranyi

Széchenyi István University, Department of Informatics, Egyetem tér 1, H-9026 Győr, Hungary, Email: baranyi.peter@sze.hu

Abstract: This paper discusses that the selection and modification of the input and the transformation space, of the Tensor Product (TP) model representation, of a given quasi–Linear Parameter Varying (qLPV) state-space model, has an influence on the feasibility regions of the Linear Matrix Inequality (LMI) based control design techniques. Moreover, three factors affect the feasibility regions of the LMI-based control design:
The manipulation of the position of vertices
The number of the inputs of the TS fuzzy model representation
Modifying the transformation space
The proof is based on a complex control design example, where the impact of the above factors can be clearly demonstrated. Furthermore, the paper presents that the maximal and minimal parameter space of the controller depends on factors (i) and (ii). The aim of the feasibility test is to show that there exists a solution for LMIs or not, considering these factors. The example is based on the academic Translational Oscillator with Rotational Actuator (TORA) system. Then, the TP model transformation-based framework is used to vary the input space of the TP model representation. In addition, the paper gives a very decisive conclusion that the design technique may be sensitive for the input space of the TS fuzzy model, hence it is necessary to consider the number of inputs, the transformation space and the gains defined on the inputs when a TP model is generated to achieve the best solution to the control purposes. All in all, this paper investigates the effect of input space

modification of the TP model representation of a given qLPV state-space model on the feasibility regions of LMI-based controller design methods.

Keywords: TP model transformation; LMI-based controller; input space; transformation space; qLPV model

1 Introduction

Three factors influence the feasibility regions of LMIs:

(i) The number and the non-linearity of the inputs

(ii) The manipulation of the position of vertices

(iii) Variations of the transformation space

The proof is based on a state feedback control design example, where these factors can be clearly detected. Accordingly, the TP model transformation is applied to generate various alternative TP models to represent the qLPV models. These alternatives have different number of inputs. Furthermore, different nonlinear gains are defined on these inputs to decrease the rank of the TP model. As a consequence, different controllers are resulted to the given qLPV model using the same Parallel Distributed Compensation (PDC) based LMI design technique, since different TP model representations have been used. This leads to a question that which one of the resulted controllers is the best one. This results in a very significant point that the design technique is sensitive to the TP model input space. Therefore, it is necessary to consider the number of inputs, the transformation space and the gains defined on the inputs when a TP model is generated to achieve the best solution to our purposes. Although it sounds quite obvious, most of the papers about TP model-based control design, do not consider this fact. Without checking the proposed method on various alternative TP models, we cannot conclude if the design method is efficient essentially or not.

Previous investigations on the topic of TP model transformation focuses on internal parameters, such as the number of antecedents and consequent fuzzy sets. On the other hand, the current article focuses on the external parameters, the number of inputs and their non-linearity. The present paper examines that the TP model may have different input space and transformation space. The suggested extension is capable of varying the number of inputs and decreasing the non-linearity and transforming the input dimensions.

The effect of the vertices of the TP model representation was examined in paper [1] via the NATA model of the three degrees of freedom Aeroelastic Wing Section, where the feasibility regions of the LMI-based controller design was influenced by the factors: manipulating of the position of vertices, the size and complexity of the model. Paper [2] examines the statement, that the convex hull of the TP model

representation has influence on the feasibility of LMI based control design. Present paper shows the feasibility regions if the transformation space is changed.

1.1 The Novel Contribution of this Paper

In the current paper, the following system is represented by the qLPV form; the academic TORA system. The controller design method includes the TP model transformation-based design framework. This paper investigates a methodical manipulation of the TP model representation complexity via changing the input and the transformation space. Furthermore, it is presented how the feasibility regions of the LMI based design change.

1.2 Preliminaries and Related Literature

The TP model transformation was developed in [3] to derive multi-level TP model structures. In addition, the Singular Value Decomposition (SVD) based method is applied to decrease a given fuzzy rule set in [4] [5]. Further details, adaption, steps and extensions of the TP model transformation can be found in books [6] [7] and papers [8] [9].

Paper [10] presents the TP model transformation of the TORA system in the control of a nonlinear benchmark problem, where the stability analysis is executed via LMI-based controller design method in PDC framework. Thus, the global asymptotic stabilization of the TORA system is executed. Therefore, it can be seen that the control design can easily be implemented. Otherwise, LMIs can give the optimal solution of the global asymptotic stabilization. Furthermore, the convex optimization via LMI-based design method can be solved rapidly.

Paper [11] shows the state-variable feedback control design of the 2DoF model of NATA Wing Section via TP model transformation and LMI based controller design method. The aim is to derive an observer for the Wing Section to estimate the non-measurable state values, then the output feedback control is designed using the LMI-based technique. The TP transformation is resulted in a tight convex hull. In case of observer design, the tight convex hull does not lead to feasible LMI. In report [12], stabilizing the 3DoF model of NATA is presented via asymptotic stability, decay rate and constraint of the control signal. Furthermore, the convex hull manipulation is investigated.

Paper [1] presents that the vertices have influence on the feasibility regions of LMI-based control design technique via the 3DoF model of Aeroelastic Wing (NATA). Furthermore, there are two factors influencing the feasibility regions: position of the vertices, size and complexity of the TP model representation. Moreover, the CNO type weighting functions are investigated from the perspective of convex hull manipulation. This has influences on the feasibility of LMI-based solutions. Paper [2] shows that the convex hull of the TP model transformation has effect on the

feasibility of LMI based stability representation. The investigation is achieved via different convex hulls.

The novel control approaches of the TP model transformation were published in [14] [25-28]. Varying the input space of the TS fuzzy model is introduced in paper [14].

For additional important applications can be found in [15-38]. Most recent results are published in [8] [13] [14] [22-24] [39-54].

1.3 Structure of the Paper

The paper is structured as follows: Section 2 defines the notation and definitions of the TP model transformation used in this paper. The proposed extension is represented in Section 3 through the statements and proofs. Section 4 presents the applied methods, for instance varying the input space, the transformation space and adapting LMI-based controller design. Furthermore, the TP transformation of the TORA system the applied weighting functions are presented in Section 5. Section 6 deals with the feasibility analysis for TORA system. The globally asymptotic stabilization of this nonlinear system is introduced in Section 7. And finally, in Section 8 conclusions are presented.

2 Notations and Definitions

In this section the notations and definitions are presented that used in current paper. For detailed definitions, see paper [14].

- Indices: i, j, k are the upper bounds of the indices e.g., $i = 1,2, \dots, I$ and $j = 1,2, \dots, J$ and $k = 1,2, \dots, K$ or $i_n = 1,2, \dots, I_n$, where $n = 1,2, \dots N$ and I, J, K are the number of the vertices
- Scalar: $a \in \mathbb{R}$
- Vector: $\mathbf{a} \in \mathbb{R}^I$ contains elements $a_i \in \mathbb{R}$
- Matrix: $\mathbf{a} \in \mathbb{R}^{I \times J}$ contains elements $a_{i,j} \in \mathbb{R}$
- Tensor: $\mathcal{A} \in \mathbb{R}^{I \times J \times K \times \dots}$ has elements $a_{i,j,k,\dots} \in \mathbb{R}$
- Interval: $\omega \subset \mathbb{R}$ is bounded as $\omega = [\omega_{min}, \omega_{max}]$
- $\mathbb{R}^{I^N}$ is brief notation of $\mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. For instance, $\mathbb{R}^{I^N \times O^K}$ denotes $\mathbb{R}^{I_1 \times I_2 \times \dots \times I_N \times O_1 \times O_2 \times \dots \times O_K}$
- System matrix $\mathbf{S}(\mathbf{p})$ is determined with parameter $\mathbf{p} = \mathbf{p}(t) \in \Omega$

- Transformation space $\Omega = [\omega_1, \omega_1] \times [\omega_2, \omega_2] \times \ldots \times [\omega_n, \omega_n] \subset \mathbb{R}^n$ denotes a closed hypercube, hence the TP model representation is interpretable only in this space
- Weighting functions $w_n = w_n(\mathbf{p}(t))$
- CNO: Close to Normalized membership function, that means its largest value is 1 or close to 1
- X-type weighting functions: Varying the input space with reduction of non-linearity results less complex weighting functions, see in Fig. 1, where $w_n(\mathbf{p_n}(t))$ are illustrated for n dimensions

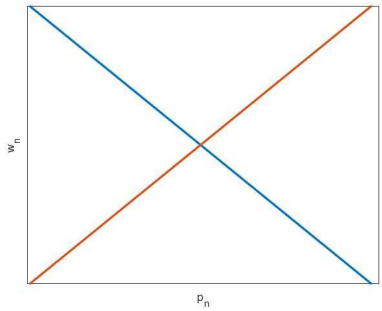


Figure 1
X-type weighting functions

- System matrix $\mathbf{S}(\mathbf{p}(t))$ is the parameter (p) dependent matrix of qLPV representation
- $\mathbf{x}(t) \in \mathbb{R}^n$ is the state vector, $\mathbf{u}(t) \in \mathbb{R}^m$ is the input vector, $\mathbf{y}(t) \in \mathbb{R}^q$ is the output vector and $\mathbf{A}_i(t) \in \mathbb{R}^{n \times n}$, $\mathbf{B}_i(t) \in \mathbb{R}^{n \times m}$

Definition 1: TP structure. Tensors can be given as a tensor product such as

$$\mathcal{S} = \mathcal{B} \bigotimes_{n=1}^{N} \mathbf{U}_n \tag{1}$$

where $\mathcal{B} \in \mathbb{R}^{I^N \times O^K}$ termed as core tensor and $\mathbf{U}_n \in \mathbb{R}^{M_n \times I_n}$ are the weighting matrices.

Definition 2: TP model. The TP model is a continuous variant of the TP structure. Here, instead of weighting vectors the weighting functions are as follow:

$$\mathcal{S}(\mathbf{p}) = \mathcal{B} \bigotimes_{n=1}^{N} [w_{n,1}(p_n) \quad w_{n,2}(p_n) \quad \ldots \quad w_{n,I_n}(p_n)] \tag{2}$$

that is $\mathcal{S}(\mathbf{p}) = \mathcal{B} \bigotimes_{n=1}^{N} \mathbf{w}_n(p_n)$ (3)

where the vector of the weighting function is $\mathbf{w}_n(p_n) = [w_{n,1}(p_n) \quad w_{n,2}(p_n) \quad \dots \quad w_{n,I_n}(p_n)] \in \mathbb{R}^{I_n}$ and $\mathcal{S}(\mathbf{p}) \in \mathbb{R}^{O^K}$, $\mathbf{p} \in \mathbb{R}^N$ and core tensor $\mathcal{B} \in \mathbb{R}^{I^N \times O^K}$ contains the vertices $\mathcal{s}_{i_1,i_2,\dots,i_N} \in \mathbb{R}^{O^K}$.

To understand the basic steps of varying the input space, it is necessary to view paper [14], that presents the hyper rectangular grid and the difference between TS fuzzy model and TP model in detail.

3 Statements and Proofs

Present section explains the statements about the TP model representation of a given qLPV model and the feasibility regions influencing the LMI-based controller design method. Furthermore, it is shown how the behavior of the LMI-based control is changed via modifying the transformation space Ω to Ω_{min} and Ω_{max} and via selecting the input space. Control design of TP models of the academic TORA system is executed by LMI. The statements are as follows:

Statement 1. Varying the input space and the non-linearity of inputs of the TP model has influence on the feasibility of the LMI-based design. Hence, the complexity of the model is reduced, but the controllers are different in each example. The proof is based on that the position of the vertices is modified (see in paper [14]).

Statement 2. It is possible to modify the transformation space Ω provided that the LMI is still feasible. Thus, it is shown how the behaviour of the LMI-based controller changes. Changing the transformation space does not always result in a feasible solution for LMI.

Statement 3. Decreasing the non-linearity through varying the input space and changing the transformation space results in increasing feasibility regions.

The proofs are based on a complex control design example of nonlinear dynamic qLPV system. The investigations follow these key points:

1) Control design and all investigations are executed on the nonlinear models.

2) Selecting the input space is based on the non-linearity of the model examples. These non-linearities belong to the original state-space representation of qLPV models. The aim is to reduce or remove these non-linearities of the different TP models of TORA system.

3) Modification of the transformation space is based on the fact, that LMI is influenced by many factors. These modifications belong to the original transformation spaces of the TP models.

Hence, the proofs show through the statements, which TP model results the better controller depending on varying the input and transformation space.

4 Applied Methods

In this section, the applied methods are presented. Manipulation of the input space is a new extension of the TP model transformation, that is discussed in paper [14]. The input space can be changed in such a way that the new inputs are functions. Thereby, the suggested method can reduce or remove the non-linearity from the TP models. However, paper [14] does not include the analysis of transformation space, so this is a novel contribution of the current investigations.

4.1 How to Vary the Input Space of the TP Model Transformation

Consider the qLPV state-space representation:

$$\begin{bmatrix}\dot{\mathbf{x}}(t)\\ \mathbf{y}(t)\end{bmatrix} = \mathbf{S}(\mathbf{p}(t))\begin{bmatrix}\mathbf{x}(t)\\ \mathbf{u}(t)\end{bmatrix} \tag{4}$$

where $\mathbf{x}(t)$, $\mathbf{u}(t)$ and $\mathbf{y}(t)$ are the state, input and output vectors, $\mathbf{S}(\mathbf{p}(t))$ is the system matrix respectively. Thus, the aim is to control this system by reducing non-linearity.

The first step is to determine the Higher Order Singular Value Decomposition (HOSVD) based canonical form for the above qLPV model [6]. The model is based on the number of inputs and the position of the vertices. Furthermore, the LMI based control design requires a convex TP model structure of the system matrix $\mathbf{S}(\mathbf{p}(t))$. In the present paper, the examples follow the current TP based model structure:

$$\begin{pmatrix}\dot{\mathbf{x}}\\ \mathbf{y}\end{pmatrix} = \left(\mathcal{S} \bigboxtimes_{n=1}^{N} \mathbf{w}_n(p_n)\right)\begin{pmatrix}\mathbf{x}\\ \mathbf{u}\end{pmatrix} \tag{5}$$

Paper [14] shows more detailed research about the new extension of the TP model transformation. Therefore, this extension is capable to transform an alternative input space. Hence, consider the following function:

$$\mathcal{S}(\mathbf{p}) \in \mathbb{R}^{O^K},\ \mathbf{p} \in \Omega^p \subset \mathbb{R}^N \tag{6}$$

then the TP model transformation for all $\mathbf{p}$ is:

$$\mathcal{S}(\mathbf{p}) = \mathcal{S} \bigboxtimes_{n=1}^{N} \mathbf{w}_n(p_n) \tag{7}$$

The alternative input space is $\mathbf{b} \in \Omega^b \subset \mathbb{R}^M$, where the relation between $\mathbf{p}$ and $\mathbf{b}$ is defined. Thus, the expected result for all $\mathbf{p}$ is:

$$\mathcal{S}(\mathbf{p}) = \mathcal{T}(\mathbf{b}) = \mathcal{A} \bigboxtimes_{m=1}^{M} \mathbf{v}_m(b_m) \tag{8}$$

where the input space $\mathbf{p} \in \Omega^p$ is replaced by $\mathbf{b} \in \Omega^b$, $\mathcal{A}$ is the core tenzor and $\mathbf{v}_m$ is the weighting function. For more detailed presentation see in paper [14].

4.2 Modification of the Transformation Space

In addition to vary the input space, present paper focuses on the investigations of extended transformation space. Ω is determined by a hyper-rectangular space, where the weighting functions are defined. Thus, the TP model representation is interpretable only in this space, see [6]. Moreover, for the examination, the transformation space to Ω_{min} and Ω_{max} is extended. These steps influence the feasibility of the LMI-based controller design method.

Current investigation is based on paper [14]. Thus, the values of the applied transformation spaces Ω for each TP models are the same, as in paper [14]. In the present paper, it is proved that varying the transformation space is possible if the LMI is feasible.

5 TP Model Transformation of TORA

In this section, the academic TORA system is investigated which was a key example to test the first variants of TP model transformation [6].

Previous research through this example [18] presented that TP model transformation and LMI-based control design approach can readily be accomplished independently of the given problem. Therefore, no analytical derivation is required, thus the controller designing is less time consuming. Those analyses focused primarily on the shape of the fuzzy antecedent sets, hence the convex hull defined by the vertices, and the number of fuzzy rules.

In the present section, we focus on the number of inputs as well as in [14]. In this example, the same 137×137 grid density and CNO type weighting functions are used. See the MATLAB code for the CNO type functions in paper [14].

5.1 qLPV Model of the TORA

The applied parameters and equations of motion are given in [6] [14] [18]. Thus, the state-space variables of TORA system are selected as:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} = \begin{bmatrix} \xi(t) \\ \dot{\xi}(t) \\ \theta(t) \\ \dot{\theta}(t) \end{bmatrix} \tag{9}$$

where, the system matrix $\mathbf{S}(\mathbf{p}(t))$ of the model takes form of:

$$\mathbf{A}(\mathbf{p}) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{-1}{f(x_3(t))} & 0 & 0 & \frac{\rho x_4(t) sin(x_3(t))}{f(x_3(t))} \\ 0 & 0 & 0 & 1 \\ \frac{\rho cos(x_3(t))}{f(x_3(t))} & 0 & 0 & \frac{-\rho^2 x_4(t) sin(x_3(t)) cos(x_3(t))}{f(x_3(t))} \end{bmatrix} \tag{10}$$

$$\mathbf{B}(\mathbf{p}) = \begin{bmatrix} 0 \\ \frac{-\rho cos(x_3(t))}{f(x_3(t))} \\ 0 \\ \frac{1}{f(x_3(t))} \end{bmatrix} \tag{11}$$

where

$$f(x_3(t)) = 1 - \rho^2 cos^2(x_3(t)) \tag{12}$$

and $p_1(t) = x_3(t)$ and $p_2(t) = x_4(t)$. The system matrix $\mathbf{S}(\mathbf{p}(t))$ is:

$$\mathbf{S}(\mathbf{p}(t)) = [\mathbf{A}(\mathbf{p}(t)) \quad \mathbf{B}(\mathbf{p}(t))] \tag{13}$$

where the elements of matrix $\mathbf{A}(\mathbf{p}(t))$ and $\mathbf{B}(\mathbf{p}(t))$ are defined numerically by HOSVD method and CNO type weighting functions into vertices.

5.2 TP Model 1

This subsection presents the previous version of the TP model transformation (7). Therefore let $p_1(t) = x_3(t)$ and $p_2(t) = x_4(t)$. The transformation space is $\Omega = [-0.8,0.8] \times [-0.8,0.8]$. The resulted TP model representation is:

$$\mathbf{S}(\mathbf{p}(t)) = \mathcal{S} \boxtimes_{n=1}^{2} \mathbf{w}_n(p_n(t)) \tag{14}$$

thus, $w_{1,i}(p_1)$ CNO type weighting functions in parameter space Ω are shown in Fig. 2. For parameter $p_2(t)$, the HOSVD results $w_{2,1}$ and $w_{2,2}$ in the X-type weighting function. The number of the resulting linear time invariant (LTI) systems is $5 \times 2 = 10$.

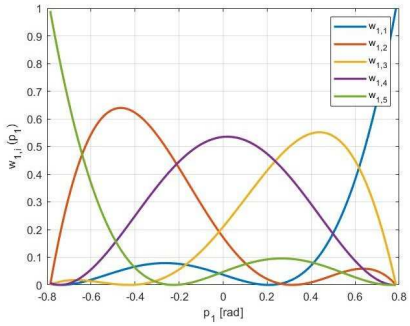


Figure 2
CNO type weighting functions of TP model 1

5.3 TP Model 2

The TP model 1 has minimal number of weighting functions. In order to decrease the non-linearity of the model, the parameter space is defined as $p_1(t) = x_3(t)$, $p_2(t) = x_4(t)$, $p_3(t) = \frac{1}{f(x_3(t))}$, where a new dimension is introduced. Thus, the transformation space is $\Omega = [-0.8,0.8] \times [-0.8,0.8] \times [1,1.05]$. Applying the HOSVD on the qLPV model 2, results in $4 \times 2 \times 2 = 16$ LTI systems. Therefore, the second TP model form is:

$$\mathbf{S}(\mathbf{p}(t)) = \mathbf{T}(\mathbf{p}(t)) = \mathcal{A} \boxtimes_{n=1}^{3} \mathbf{w}_n(p_n(t)) \tag{15}$$

The membership functions of $p_1(t)$ are illustrated in Fig. 3. For $p_2(t)$ and $p_3(t)$, weighting functions $w_{2,1}$,$w_{2,2}$ and $w_{3,1}$, $w_{3,2}$ are depicted on the Fig. 1 X-type. Furthermore, the complexity of the weighting functions is reduced. However, the non-linearity can be further reduced, see in TP model 3.

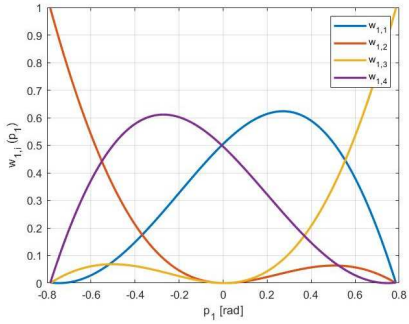


Figure 3
CNO type weighting functions of TP model 2

5.4 TP Model 3

In order to further decrease the complexity of the first dimension, the new input parameters are $p_1(t) = sin(x_3(t))$, $p_2(t) = x_4(t)$ and $p_3(t) = cos(x_3(t))$. Then the space $\Omega = [-0.8,0.8] \times [-0.8,0.8] \times [0,0.8]$ is defined. The resulting number of the LTI systems is $2 \times 2 \times 3 = 12$, thus the TP model structure is:

$$\mathbf{S}(\mathbf{p}(t)) = \mathbf{T}(\mathbf{p}(t)) = \mathcal{A} \boxtimes_{n=1}^{3} \mathbf{w}_n(p_n(t)) \tag{16}$$

The weighting functions of $p_3(t)$ are shown in Fig. 4. For $p_1(t)$ and $p_2(t)$, weighting functions $w_{1,1}$,$w_{1,2}$ and $w_{2,1}$, $w_{2,2}$ are depicted on the Fig. 1 X-type. The rank of the first dimension is minimised, in the third dimension there are 3 weighting functions in contrast to TP model 2. Consequently, the resulted model is less complex than the previous two examples.

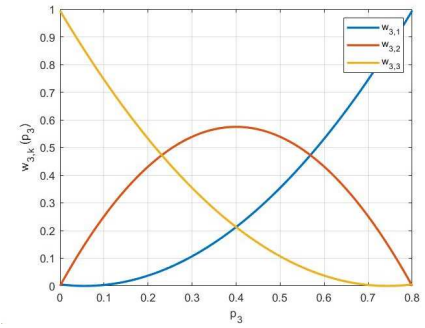


Figure 4
CNO type weighting functions of TP model 3

5.5 TP Model 4

Let is the parameter space $p_1(t) = x_4(t)sin(x_3(t))$, $p_2(t) = \frac{1}{f(x_3(t))}$ and $p_3(t) = \rho cos(x_3(t))$, and the transformation space $\Omega = [0{,}1.6] \times [1{,}1.05] \times [0{,}0.2]$. Executing the proposed TP model transformation, there are $2 \times 2 \times 2 = 8$ LTI systems:

$$\mathbf{S}(\mathbf{p}(t)) = \mathbf{T}(\mathbf{p}(t)) = \mathcal{A} \underset{n=1}{\overset{3}{\boxtimes}} \mathbf{w}_n(p_n(t)) \tag{17}$$

The membership functions are depicted in Fig. 1, where the weighting functions $w_{1,1}$, $w_{1,2}$, $w_{2,1}$, $w_{2,2}$ and $w_{3,1}$, $w_{3,2}$ for parameters $p_1(t)$, $p_2(t)$, and $p_3(t)$ are illustrated. Compared to the previous TP model 1,2,3 examples, this is the simplest model.

5.6 TP Model 5

Let us define the parameter space as $p_1(t) = \frac{x_4(t)sin(x_3(t))}{f(x_3(t))}$, $p_2(t) = \frac{1}{f(x_3(t))}$ and $p_3(t) = cos(x_3(t))$. and $\Omega = [0{,}1.6] \times [1{,}1.05] \times [0{,}1]$. The number of the LTI systems are $2 \times 2 \times 2 = 8$ again:

$$\mathbf{S}(\mathbf{p}(t)) = \mathbf{T}(\mathbf{p}(t)) = \mathcal{A} \underset{n=1}{\overset{3}{\boxtimes}} \mathbf{w}_n(p_n(t)) \tag{18}$$

The membership functions are given in X-type for all parameters $p(t)$. This is similar to the previous example, but the consequences are different from it.

6 Feasibility Analysis

The feasibility test checks whether there exists a solution for LMI or not. Consider the following solver for LMI feasibility problems: $\mathbf{L}(x) < \mathbf{R}(x)$, where $\mathbf{R}$ is the feasibility radius. The solver minimizes t subject to $\mathbf{L}(x) < \mathbf{R}(x) + t\mathbf{I}$. Thus, the best value of t should be negative for feasibility.

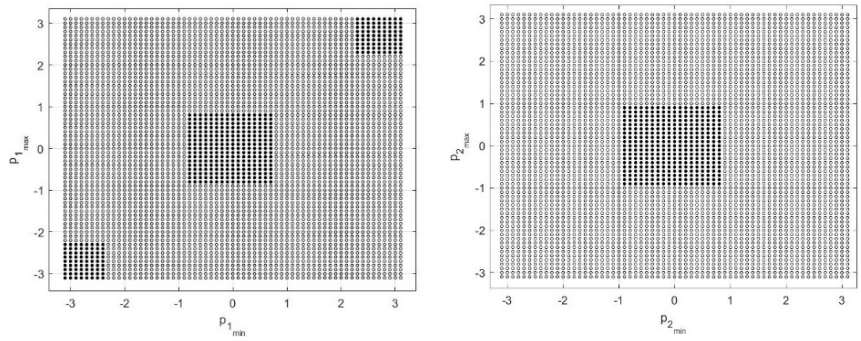


Figure 5
Feasibility regions of TP model 1

The feasibility check for transformation space is investigated only on interval $[-180°, 180°]$ for angles. For all figures; x-axis illustrates p_{min} and y-axis shows p_{max}, illustrations present that LMI is feasible or not in the different dimensions. Therefore, figures show the feasible (bold dotted) and non-feasible regions of the LMI through changing transformation space. In this section the most significant TP models and figures are presented.

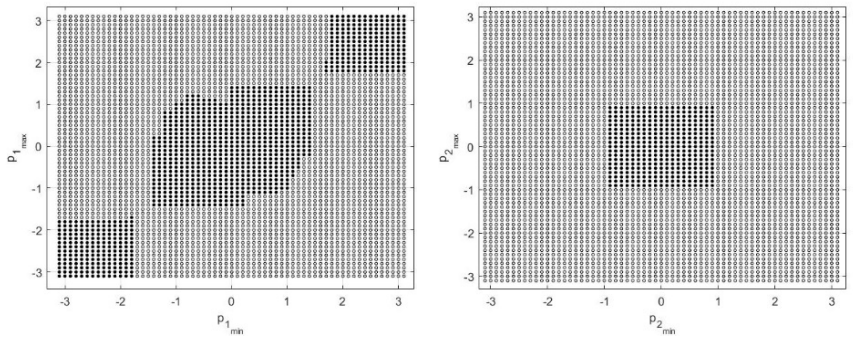


Figure 6
Feasibility regions of TP model 3 for p_1 and p_2

Varying the transformation space Ω has influence on the feasibility regions of LMIs. In the followings the feasibility of LMI based controller for TORA TP models are presented.

In the present example, there are two dimensions in the transformation space $\Omega = [-0.8,0.8] \times [-0.8,0.8]$. The parameters are $p_1(t) = x_3(t)$, $p_2(t) = x_4(t)$. Moreover, Fig. 5 presents two cases. In the first case, the first dimension p_1 of the transformation space is changed, and second dimension p_2 is fixed on interval $[-,-] \times [-0.8,0.8]$. Fig. 5 shows another case, that first dimension p_1 is fixed on $[-0.8,0.8] \times [-,-]$ and second dimension p_2 is changed. Then, it can be seen that LMI is feasible or not, in the dimensions of TORA TP model 1.

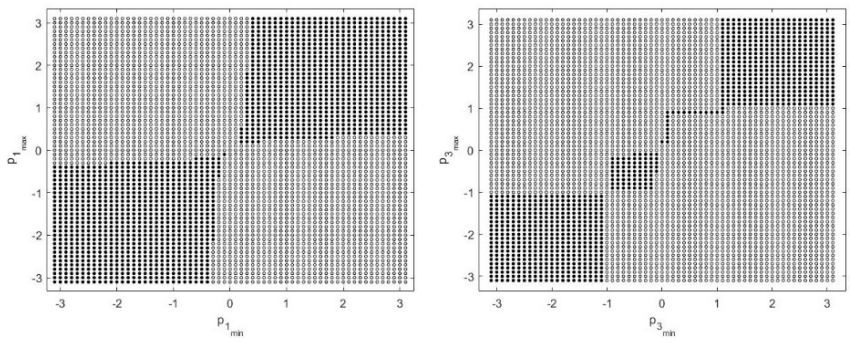


Figure 7
Feasibility regions of TP model 4 for p_1 and p_3

In case of TORA TP model 2, there are three dimensions; $\Omega = [-0.8,0.8] \times [-0.8,0.8] \times [1,1.05]$, where $p_1(t) = x_3(t)$, $p_2(t) = x_4(t)$, $p_3(t) = \frac{1}{f(x_3(t))}$. As before, two cases are considered; each of the two dimensions is fixed separately. Thus, the results for changing dimensions p_1 and p_2 are the same as in Fig. 5.

The feasibility tests for TORA TP model 3 are illustrated on Fig. 6, where the transformation space $\Omega = [-0.8,0.8] \times [-0.8,0.8] \times [0,0.8]$, and $p_1(t) = sin(x_3(t))$, $p_2(t) = x_4(t)$ and $p_3(t) = cos(x_3(t))$. Figures shows that dimension p_1 is changed and, the others are fixed on interval $[-,-] \times [-0.8,0.8] \times [0,0.8]$, then dimension p_2 is changed on interval $[-0.8,0.8] \times [-,-] \times [0,0.8]$. It can be seen, that the feasibility regions are increased compared to TP model 1 and 2.

Fig. 7 present the feasibility regions of TORA TP model 4. In this case, the transformation space is $\Omega = [0,1.6] \times [1,1.05] \times [0,0.2]$, and the parameter space is $p_1(t) = x_4(t)sin(x_3(t))$, $p_2(t) = \frac{1}{f(x_3(t))}$ and $p_3(t) = \rho cos(x_3(t))$. Two cases are examined again. Thus, it can be seen the feasibility regions if dimension p_1 is changed on interval $[-,-] \times [1,1.05] \times [0,0.2]$ and if dimension p_3 is changed on interval $[0,1.6] \times [1,1.05] \times [-,-]$. The feasibility regions are larger, than the previous TP models.

Fig. 8 illustrates the feasibility regions of TORA TP model 5. The transformation space is $\Omega = [0,1.6] \times [1,1.05] \times [0,1]$ and the parameters are $p_1(t) = \frac{x_4(t)sin(x_3(t))}{f(x_3(t))}$, $p_2(t) = \frac{1}{f(x_3(t))}$ and $p_3(t) = cos(x_3(t))$. So, the figures show two cases. In the first case, dimension p_1 is changed on interval $[-,-] \times [1,1.05] \times [0,1]$ and for p_2 and p_3 are fixed. In the second case, dimension p_3 is changed on interval $[0,1.6] \times [1,1.05] \times [-,-]$, and p_1 and p_2 are fixed on this interval. It can be seen that the feasibility regions are larger, than the previous four TP models.

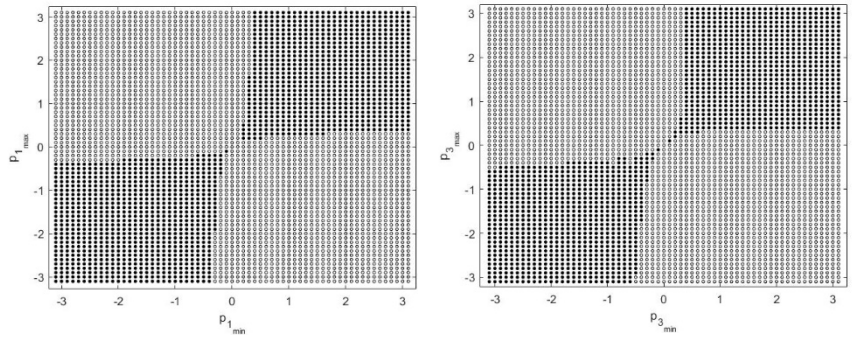


Figure 8
Feasibility regions of TP model 5 for p_1 and p_3

Consequently, the figures present well that if the nonlinear behavior of the TP models are decreased, the LMI feasibility regions will increase.

7 Globally Asymptotic Stabilization

In this section, all of the examined TP models are illustrated. The results of the control design of the TORA system are presented via LMI based stabilization method, through state feedback control. The LMIs are feasible. The LMI solver can find the stabilizing controller using the following LMI conditions [10]:

$$-\mathbf{X}\mathbf{A}_i^T - \mathbf{A}_i\mathbf{X} + \mathbf{M}_i^T\mathbf{B}_i^T + \mathbf{B}_i\mathbf{M}_i \succ 0 \tag{19}$$

$$-\mathbf{X}\mathbf{A}_i^T - \mathbf{A}_i\mathbf{X} - \mathbf{X}\mathbf{A}_j^T - \mathbf{A}_j\mathbf{X} + \mathbf{M}_j^T\mathbf{B}_i^T + \mathbf{B}_i\mathbf{M}_j$$

$$+\mathbf{M}_i^T\mathbf{B}_j^T + \mathbf{B}_j\mathbf{M}_i \succeq 0 \tag{20}$$

$$\mathbf{k}_i = \mathbf{M}_i\mathbf{X}^{-1} \tag{21}$$

where $i = 1, \cdots, I$ and $j = I+1, \cdots, I$, I is the total number of the LTI vertex systems. Using these LMI conditions, present paper shows a stable controller design task. Matrices $\mathbf{X}$ and $\mathbf{M}$ can be found by convex optimization methods involving LMIs.

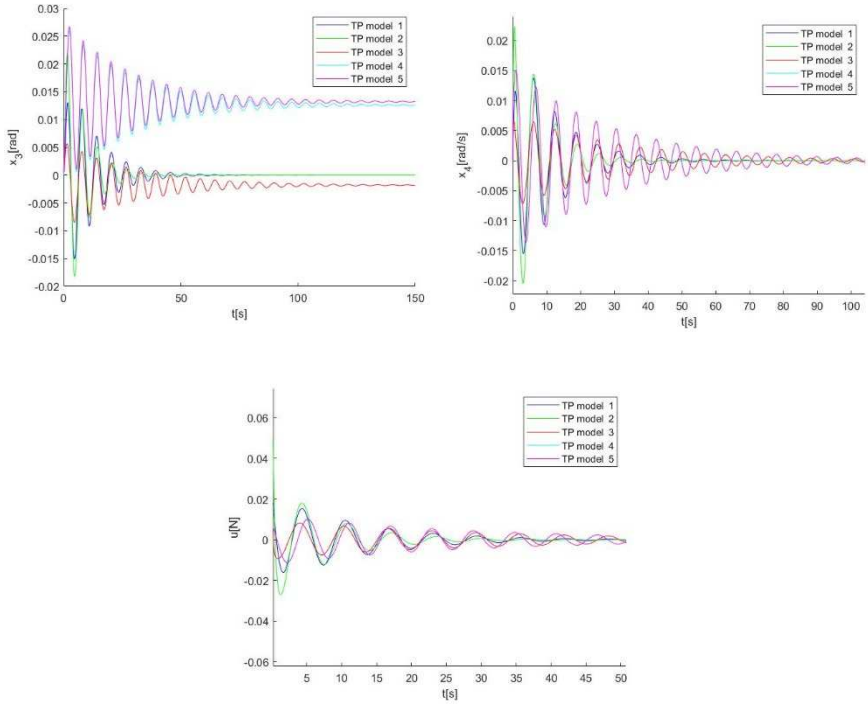


Figure 9
State variables x_3, x_4 and control signal u of TORA

The resulted LTI vertex systems defined by TP transformation are substituted into the above LMI conditions.

The overall TP controller is:

$$\mathbf{u}(t) = -(\sum_{i=1}^{I} w_i(\mathbf{p}(t))\mathbf{k}_i)))\mathbf{x}(t) = -\mathbf{k}^T\mathbf{x}(t) \tag{22}$$

where the same membership functions are applied as in the TP model examples. It can be seen that the various inputs of the TS fuzzy models influence the feasibility of the LMI based controller design.

Angular position θ, and angular speed $\dot{\theta}$ of TP model 1, 2, 3, 4, 5 examples and control signal u are illustrated in Fig. 9, where the overall initial conditions are $[\xi, \dot{\xi}, \theta, \dot{\theta}]^T = [0.023,0,0,0]^T$. Notice that the TP model 2, results in the best controller performance. In example TP model 2, a new dimension is introduced and the number of LTI systems are $4 \times 2 \times 2 = 16$. Although, the complexity of the weighting functions is still exists, decreasing the non-linearity results in a better controller.

It can be seen that control signal u is stabilized after 50 seconds. Therefore, the simulation results show that the system reaches the steady state quicker and the

complexity of the TP model 2 is less than the other ones. Furthermore, varying the input space, it results in different TP models, as shown in Fig. 3. Number of the dimensions and complexity of the weighting function have influence on the feasibility of LMI based control design method. Consequently, the TP model 2 is the better model.

Conclusions

This work shows the proposed extension of the TP model representation, where it is capable of generating an alternative input space. This step has influence on non-linearity and complexity of the model. In this paper, the investigation is based on the TORA system. Moreover, modifying the transformation space affects the LMI-based control design methods. Consequently, the controller can be sensitive for varying the input space and transformation space. Decreasing the non-linearity via changing the input space, results in increasing feasibility regions varying the transformation space. Furthermore, the new properties improve the capability of the TP model in achieving the best possible solution of the control design methods.

References

[1] A. Szollosi, P. Baranyi: Influence of the tensor product model representation of qLPV models on the feasibility of linear matrix inequality, Asian Journal of Control, Vol. 18, No. 4, pp. 1328-1342, 2016

[2] A. Szollosi, P. Baranyi: Influence of the tensor product model representation of qLPV models on the feasibility of linear matrix based stability analysis, Asian Journal of Control, Vol. 20, No. 1, pp. 531-547, 2018

[3] Y. Yam: Fuzzy approximation via grid point sampling and singular value decomposition, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 27, No. 6, pp. 933-951, 1997

[4] Y. Yam, P. Baranyi, and C. Yamg: Reduction of fuzzy rule base via singular value decomposition, IEEE Transactions on Fuzzy Systems, Vol. 7, No. 2, pp. 120-132, April 1999

[5] P. Baranyi and Y. Yam: Fuzzy rule base reduction, Fuzzy IF-THEN Rules in Computational Intelligence: Theory and Applications, D. Ruan and E. E. Kerre, Eds. Kluwer, ch. 7, pp. 135-160, 2000

[6] P. Baranyi, Y. Yam, and P. Várlaki: Tensor Product Model Transformation in Polytopic Model Based Control, ser. Automation and Control Engineering. CRC Press, Taylor & Frances Group, 2017, ISBN 9781138077782 - CAT K34341

[7] P. Baranyi: TP-Model Transformation Based Control Design Frameworks, ser. Control Engineering. Springer book, 2016, ISBN 978-3-319-19605-3

[8] P. Baranyi: Extracting lpv and qlpv structures from state-space functions: A TP model transformation based framework, IEEE Transactions on Fuzzy Systems, Vol. 28, No. 3, pp. 499-509, 2019

[9] P. Baranyi: TP model transformation as a way to LMI-based controller design, IEEE Transactions on Industrial Electronics, Vol. 51, No. 2, pp. 387-400, 2004

[10] K. Tanaka and H. Wang: Fuzzy control systems design and analysis: a linear matrix inequality approach, John Wiley and Sons, Inc., U.S.A., 2001

[11] P. Baranyi: Tensor product model based control of two dimensional aeroelastic system, Journal of Guidance, Control and Dynamics, Vol. 29, No. 2, pp. 391-400, 2006

[12] B. Takarics and P. Baranyi: Tensor-product-model-based control of a three degrees- of-freedom aeroelastic model, Journal of Guidance, Control, and Dynamics, Vol. 36, No. 5, pp. 1527-1533, 2013

[13] M. Kuczmann: TP-based Control of the Crane Model, 2019 1st IEEE International Conference on Gridding and Polytope Based Modeling and Control (GPMC)

[14] P. Baranyi: How to vary the input space of a ts fuzzy model: a TP model transformation based approach, IEEE Transactions on Fuzzy Systems, Vol. 30, No. 2, pp. 345-356, 2022

[15] V. C. d. S. Campos, L. M. S. Vianna, M. F. Braga: A tensor product model transformation approach to the discretization of uncertain linear systems, Acta Polytechnica Hungarica, Vol. 15, No. 3, 2018

[16] A. Szollosi and P. Baranyi: Improved control performance of the 3-DoF aeroelastic wing section: a TP model based 2D parametric control performance optimization, Asian Journal of Control, Vol. 19, No. 2, pp. 450-466, 2016

[17] P. Korondi: Sector sliding mode design based on tensor product model transformation, Proceedings of the 11th IEEE International Conference on Itelligent Engineering Systems (INES) , pp. 253-258, 2007

[18] Z. Petres, B. Reskó, and P. Baranyi: Tp model transformation based control of the TORA system, Production Systems and Information Engineering, Vol. 2, pp. 159-175, 2004

[19] B. Takarics and Y. Yam: Robust grid point-based control design for LPV systems via unified TP transformation, in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2015, pp. 2626-2631

[20] A. Rövid, L. Szeidl, and P. Várlaki, “On tensor-product model based representation of neural networks,” in Proceedings of the 15th IEEE International Conference on Itelligent Engineering Systems (INES), 2011, pp. 69-72

[21] B. Takarics, A. Szollosi, and B. Vanek: Tensor product type polytopic LPV modeling of aeroelastic aircraft, in IEEE Aerospace Conference, 2018, pp. 1-10

[22] A. Wéber, M. Kuczmann: TP Transformation Based Controller and Observer Design of the Inverted Pendulum, PRZEGLAD ELEKTROTECHNICZNY, ISSN 0033-2097, R. 98 NR 10/2022, pp. 34-39, 2022

[23] A. Wéber, M. Kuczmann: TP Transformation of the Inverted Pendulum, 2019 1st IEEE International Conference on Gridding and Polytope Based Modeling and Control (GPMC), pp. 13-18

[24] A. Wéber, M. Kuczmann: Study of the TP transformation via the inverted pendulum example, Acta Technica Jaurinensis, Vol. 14, No. 1, pp. 1-23, 2021

[25] X. Liu, Y. Yu, Z. Li, and H. H. C. Lu: Polytopic H_∞ filter design and relaxation for nonlinear systems via tensor product technique, Signal Processing, Elsevier, Vol. 127, pp. 191-205, 2016

[26] G. Zhao, D. Wang, and Z. Song: A novel tensor product model transformation-based adaptive variable universe of discourse controller, Journal of the Franklin Institute, Elsevier, Vol. 353, No. 17, 2016

[27] V. C. S. Campos, F. O. Souza, L. A. B. Torres, and R. M. Palhares: New stability conditions based on piecewise fuzzy lyapunov functions and tensor product transformations, IEEE Transactions on Fuzzy Systems, Vol. 21, No. 4, pp. 784-760, 2013

[28] S. Kuntanapreeda: Tensor product model transformation based control and synchronization of a class of fractional-order chaotic systems, Asian Journal of Control, Vol. 17, No. 2, pp. 371-3, 2015

[29] G. Zhao, H. Li, and Z. Song: Tensor product model transformation based decoupled terminal sliding mode control, International Journal of Systems Science, Taylor & Frances, Vol. 47, No. 8, pp. 1791-1803, 2014

[30] P. Korondi: Tensor product model transformation-based sliding surface design, Acta Polytechnica Hungarica, Vol. 3, No. 4, pp. 23-35, 2006

[31] T. Wang, W. Xie, G. Liu, and Y. Zhao: Quasi-min-max model predictive control for image-based visual servoing with tensor product model transformation, Asian Journal of Control, Vol. 17, No. 2, pp. 402-416, 2015

[32] J. Matusko, S. Iles, F. Kolonic, and V. Lesic: Control of 3D tower crane based on tensor product model transformation with neural friction compensation, Asian Journal of Control, Vol. 17, No. 2, pp. 443-458, 2015

[33] S. Campos, V. Costa, L. A. B. Torres, and R. M. Palhares: Revisiting the TP model transformation: Interpolation and rule reduction, Asian Journal of Control, Vol. 17, No. 2, pp. 392-401, 2015

[34] F. Kolonic, A. Poljugan, and I. Petrovic: Tensor product model transformation-based controller design for gantry crane control system-an application approach, Acta Polytechnica Hungarica, Vol. 3, No. 4, pp. 95-112, 2006

[35] P. Grof and Y. Yam: Furuta pendulum-a tensor product model-based design approach case study, in Proceedings of the IEEE International Conferenc on Systems, Man, and Cybernetics (SMC), 2015, pp. 2620-2625

[36] S. Iles, J. Matusko, and F. Kolonic: Tensor product transformation based speed control of permanent magnet synchronous motor drives, in Proceedngs of the 17th International Conference on Electrical Drives and Power Electronics (EDPE) (5th Joint Slovak-Croatian Conference), 2011

[37] S. Iles, J. Matusko, and F. Kolonic: TP transformation based control of rotary pendulum, in Proceedings of the 34th IEEE International Convention, 2011, pp. 833-839

[38] S. Iles, F. Kolonic, and J. Matusko: Linear matrix inequalities based H_∞ control of gantry crane using tensor product transformation, in Proceedings of the 18th International Conference on Process Control, 2011

[39] Z. Németh, M. Kuczmann: Tensor Product Transformation based Modelling of Induction Machine, 2019 1st IEEE International Conference on Gridding and Polytope Based Modeling and Control (GPMC)

[40] H. Bouzaouache: Tensor product-based model transformation and optimal controller design for high order nonlinear singularly perturbed systems, Asian Journal of Control, 2019

[41] T. Wang, Y. Bi, T. Hou, B. Liu, and J. Cui: An efficient tp model transformation algorithm for robust visual servoing in the presence of uncertain data, Acta Polytechnica Hungarica, Vol. 17, No. 6, 2020

[42] V. C. d. S. Campos, M. F. Braga, and L. Frezzatto: Analytical upper bound for the error on the discretization of uncertain linear systems by using the tensor product model transformation, Acta Polytechnica Hungarica, Vol. 17, No. 6, pp. 61-74, 2020

[43] A. Boonyaprapasorn, S. Kuntanapreeda, T. Sangpet, P. S. Ngiamsunthorn, and E. Pengwang: Biological pest control based on tensor product transformation method, Acta Polytechnica Hungarica, Vol. 17, No. 6, pp. 25-40, 2020

[44] B. Shi, F. Chang, S. Huang, G. Zhao, Y. Li: Control Design of GOOGOL's Mobile Two-Wheeled Self-Balancing Robot Based on TP Model Transformation and Non-fixed-time Step Sampling Method, IEEE 2021 40th Chinese Control Conference (CCC), pp. 1520-1525, 2021

[45] E-L. Hedrea, R-E. Precup, R-C. Roman, E. M. Petriu: Tensor product-based model transformation approach to tower crane systems modeling, Asian Journal of Control, Vol. 23, No. 3, pp. 1313-1323, 2021

[46] M. Kuczmann: Study of tensor product model alternatives, Asian Journal of Control, Vol. 23, No. 3, pp. 1249-1261, 2020

[47] B. Shi, G. Zhao, S. Huang: Local extrema refinement-based tensor product model transformation controller with problem independent sampling methods, International Journal of Control, Automation and Systems, Vol. 20, pp. 1351-1364, 2022

[48] Z. Németh, M. Kuczmann: Tensor product transformation-based modeling of an induction machine, Asian Journal of Control, Vol. 23, No. 3, pp. 1280-1289, 2020

[49] P. Várlaki, L. Palkovics, A. Rövid: On modeling and identification of empirical partially intelligible white noise processes, Asian Journal of Control, Vol. 23, No. 3, pp. 1262-1279, 2020

[50] A. B. Csapo: Cyclical inverse interpolation: An approach for the inverse interpolation of black-box models using tensor product representations, Asian Journal of Control, Vol. 23, No. 3, pp. 1301-1312, 2021

[51] A. Boonyaprapasorn, S. Kuntanapreeda, P. S. Ngiamsunthorn, and E. Pengwang, T. Sangpet: Tensor product model transformation-based control for fractional-order biological pest control systems, Asian Journal of Control, Vol. 23, No. 3, pp. 1340-1351, 2020

[52] E.-L. Hedrea, R.-E. Precup, E. M. Petriu, C.-A. B.-Dragos, C. Hedrea: Tensor product-based model transformation approach to cart position modeling and control in pendulum-cart systems, Asian Journal of Control, Vol. 23, No. 3, pp. 1238-1248, 2021

[53] E.-L. Hedrea, R.-E. Precup, R.-C. Roman, E. M. Petriu: Tensor product-based model transformation approach to tower crane systems modeling, Asian Journal of Control, Vol. 23, No. 3, pp. 1313-1323, 2021

[54] S. Iles, J. Matusko, M. Lazar: Piece-wise ellipsoidal set-based model predictive control of linear parameter varying systems with application to a tower crane, Asian Journal of Control, Vol. 23, No. 3, pp. 1324-1339, 2021