Development of a Viscosity Model and an Application, for the Filling Process Calculation in Visco-Dampers

Márk Venczel[1], Árpád Veress[1,2], Zoltán Peredy[3]

[1]Department of Aeronautics and Naval Architecture, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3, H-1111 Budapest, Hungary, e-mail: mvenczel@vrht.bme.hu

[2]Knorr-Bremse Brake Systems Ltd., R&D Center Budapest, Major utca 69, H-1119 Budapest, Hungary, e-mail: arpad.veress@knorr-bremse.com

[3]Engineering Institute, Edutus University, Stúdium tér 1, H-2800 Tatabánya, Hungary, e-mail: peredy.zoltan@edutus.hu

Abstract: The lifetime and thermal management of torsional vibration dampers, working with silicone oils, begin at the design and development phases, of this damping product, with help of properly developed material and CFD simulation models. Dynamic viscosity measurements have been carried out on AK 1 000 000 STAB silicone oil samples, with a high-precision rotational rheometer, to quantify the temperature and shear rate dependence of the mentioned silicone oil's non-Newtonian viscous behavior. Eight commonly used pseudoplastic non-Newtonian viscosity models have been selected for model parameter estimation and comparison, to develop a reliable, accurate and easy-to-implement viscosity model valid in -40°C-200°C temperature and 0 1/s-1000 1/s shear rate ranges, based on the expected operational range of the fluid. The filling process of an existing, small-sized torsional vibration damper has been thoroughly analyzed under eight different filling conditions and filling time comparison. Filling mass characteristics, maximal allowed oil inlet temperatures and maximal oil filling velocities have been determined for each case with help of 3D, transient, multiphase, coupled fluid dynamic and heat transfer calculations.

Keywords: torsional vibration damper; silicone oil; rheology; pseudoplastic viscosity model; multiphase simulation; CFD; Carreau-Yasuda model

1 Introduction

Due to the versatility and favorable physical, chemical and mechanical properties of silicone oils, they are widely used in almost all areas of life, starting from the medical, food and cosmetic industries to instrument technology, production

technology and scientific-theoretical research, up to machinery, vehicle and aerospace applications.

Within the vehicle industry, silicone oils play an essential role in damping of harmful torsional oscillations awakening on the crankshaft of high-performance internal combustion engines to extend the engines' lifetime. Viscous torsional vibration dampers (shortly visco-dampers or TVDs) will remain the primary working device in all large-scale internal combustion engines (e.g., ships) where the generation, maintenance and regulation of the electromagnetic field cannot be solved in an efficient and economical way for the application of a fluid-damper operating on the magnetorheological principle [1]. Apart from the vehicle industry utilizations, visco-dampers are preferred not only in the energetic and mining sectors, but they are also applied in the markets of combined heat and power, industrial manufacturing & processing, energy and utilities, landfill, biogas and agriculture.

Since visco-dampers are considered a multi-industry component of high importance from the safety point of view, they have a strong business and economic background. While the internal combustion engine market is estimated to be worth USD 271,508.6 million by 2026 (6.5% Compound Annual Growth Rate from 2018-2026), the visco-damper market is estimated to be USD 2360 million by 2024 (2% average growth rate between 2019 and 2024 [2]).

The durability of a visco-damper is determined by the lifetime of the silicone oil stored in it. The oil's wearing and degradation process occur not only during the operation of the damping device (vibration damping process) due to the high thermal, strong mechanical and long-lasting chemical effects, but it starts at the production phase of the damper when the oil is filled into the damper gap.

The aim of this paper is to investigate the filling process by considering the minimally allowed oil degradation by using fluid flow and thermal calculations for saving cost, time and capacity and to optimize the process also to reach the same goals during the production process. Hence, following the introduction of the visco-dampers and their damping medium as silicone oils, the second step of this report is to present the developments of a reliable and accurate silicone oil material model to be the most suitable for AK 1 000 000 STAB silicone oil in order to apply it in a 3D, transient, multiphase, coupled fluid dynamic and heat transfer calculation suitable for testing and optimizing the filling process of visco-dampers. Eight non-Newtonian viscosity models are used, tested, and compared with each other in terms of best parameter identification and accuracy. By selecting the most appropriate viscosity model, the maximum allowable filling inlet temperature for the mentioned silicone oil is also identified for the different filling inlet overpressure cases (10^5 Pa, 5×10^5 Pa, 10×10^5 Pa and 20×10^5 Pa) calculated numerically, on a small-sized visco-damper gap geometry. The filling time required to push the same amount of oil (25 g) into the damper is calculated, and the highest total temperature are also determined for each investigated filling case.

1.1 Visco-Dampers in the Vehicle Industry

The crankshaft of a high-performance internal combustion engine is loaded with axial, transverse and torsional oscillations. Among these undesirable phenomena, the most dangerous are the torsional vibrations [3], which cannot be absorbed by the crankshaft's support bearings. If the frequency of the harmful torsional oscillations appearing on the crankshaft matches with the natural frequency of the crankshaft and the driven parts, resonance develops, and fatigue fracture occurs.

To avoid the above-mentioned undesirable phenomenon and to reduce the amplitude of harmful torsional oscillations, one possibility is to protect the crankshaft with a torsional vibration damper. This damping device is installed either on the free end of the crankshaft or integrated into the flywheel. There are several types of dampers, such as frictional, spring or rubber, but the viscous version is the simplest in structure and requires the least maintenance. According to Figure 1, the visco-damper is essentially a ring-shaped closed space (housing) in which a solid ring (inertia-ring) can move freely and is guided by slide bearings. The narrow space between the housing and the inertia-ring (oil gap) is filled with high-viscosity silicone oil.

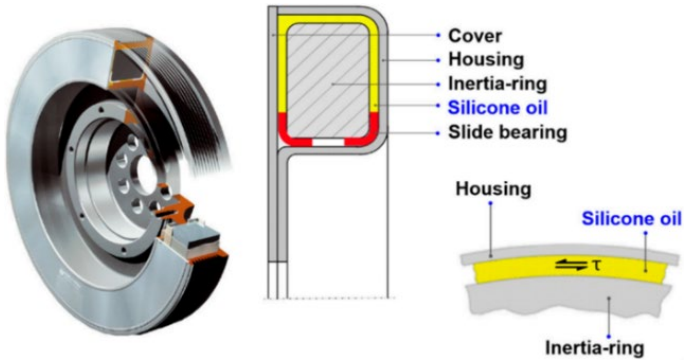


Figure 1
Structure and operation of a visco-damper [4]

The housing is attached to the crankshaft, it rotates together with the shaft, and if even a small torsional oscillation appears on the crankshaft, the housing begins a relative movement to the inertia-ring and the oil in the oil gap is sheared in the tangential direction. The sum of the tangential shear forces on the friction surfaces results in a damping effect, because rotationally the ring is accelerated, and the housing is decelerated. Since silicone oil is a non-Newtonian fluid (see details in subsection 1.2), the tangential shear resulting from the velocity difference between the housing and the inertia-ring and the temperature, during engine operation, have an impact on the viscosity of the silicone oil and on the viscous damping process taking place in the fluid. The torsional oscillations appearing on the crankshaft, are converted into thermal energy (heat) by means of friction and shear in the damper,

while the silicone oil suffers from significant mechanical, thermal and chemical wear.

There are several research projects are dealing with modelling and simulating the operational process of the visco-dampers for saving time, cost and capacity in the design and development phases.

In 2017, Pistek et al. [5] focused on the mathematical description of convolute rheological characteristics of the damping fluid of visco-dampers with the application of the traditional fourth-order Maxwell model and with the use of the general rheological model. The optimization program GAMS's CONOPT solver was used to construct a computational tool for calculating the stiffness and damping coefficients of the multi-parameter rheological model. The accuracy of their sophisticated dynamic computational model was validated by tensometric torque measurements at the crank pin.

In 2021, Homik et al. [6] proposed a thermo-hydrodynamic damper model that provides a good approximation of the operating temperature range of a visco-damper. The introduced damper model is able to consider the nature of excitation from the drive unit, geometric parameters, as well as physical, kinematic and dynamic properties of the damper. The input parameters for the model, specifically the angular velocities of the damper parts and the geometry and mass dimensions of the damper were obtained on a test bench using a six-cylinder diesel engine of the Andoria 6C107 equipped with a factory torsional vibration damper. The damper surface operating temperatures used in model verification were measured with a laser pyrometer.

These mathematical models can describe the damping process, but it is hard to investigate the inherent processes of the silicon oils in an arbitrary 3D geometry.

Concerning the silicon oils in the visco-dampers, depending on the degree and length of the fluid's periodical stress, one can talk about a temporary viscosity loss, which is a planned state of the damping process, and one can talk about a permanent viscosity loss, which is called degradation, which should be avoided or delayed, as it results in a significant decreasing in the silicone fluid’s lifetime (and at the same time the entire TVD). The wear and tear process just outlined can start not only during the engine’s operation but also in the last phase of the damper’s production process, when the device is filled with silicone oil (see more details of the filling process in Section 3).

The aspects are mentioned in the paragraph above are the reasons why it is necessary to model and numerically analyze the filling process of a TVD and identify the boundary conditions of the filling, under which the oil suffers minimal degradation from a thermal point of view in the production process. However, the first and most important step of this task is the development of a reliable and accurate silicone oil viscosity model based on rheological measurements (see subsection 2.2). But before solving the task, let us introduce the applied working medium in the next subsection.

1.2 Silicone Oils with Wide Application Ranges

Silicone polymers exist in many different forms in our daily life, such as liquids, greases, rubbers and resins. This article deals only with linear polydimethylsiloxane (PDMS), because the silicone oil (subject of this article) is also made up of this material.

PDMS is a synthetic polymeric organosilicon fluid, which do not exist in nature. Under normal (ambient) conditions, PDMS is a clear, colorless, odorless, pure and viscous liquid without detectable vapor pressure. It is considered physiologically inert, non-toxic, chemically neutral and have no marked harmful impact on living organisms in the environment [7].

Considering the silicone oils' specific viscous character and non-Newtonian behavior, they are primary working fluids in vibration damping applications. Out of the already mentioned applications in vehicle industry, the silicon oils are found in many different dampers widely used for instance in earthquake protection of civil buildings, weapons, spacecraft or in high-precision machinery and instrumentation to avoid impact-vibration damage by absorbing energy of vibrations or shocks and suppress the resonance or alleviate the shock acceleration.

As far as the thermal properties of silicone oils is concerned, they have favorable heat-transfer characteristics with a relatively high thermal stability, so PDMS fluids are perfect thermic lubricants, transformator and hydraulic fluids. They are excellent electrical insulators [8] and, unlike their carbon analogues, are non-flammable. Their temperature stability and good heat-transfer characteristics [9] make them widely used in laboratories for heating baths ('oil baths') placed on top of hotplate stirrers, as well as in freeze-dryers as refrigerants. Silicone oil is also commonly used as the working fluid in dashpots, diffusion pumps and in oil-filled heaters. In electronics they are used as a temperature-resistant, arc-resistant, corona-resistant, anticorrosive, moisture-proof and dustproof insulating medium [10] on electric motors, electrical appliances and electronic instruments. They found to be the best impregnating agent of capacitors and TV scan transformers.

In terms of silicone oils' chemical properties, they are hydrophobic, have relative low surface tension and they wet almost any surface well, thus PDMS is added to many cooking oils (as an antifoaming agent) to prevent oil splatter during the cooking process. PDMS is widely used in anti-dust and anti-oil fluids as additive. In aerospace it is used as aircraft hydraulic oil thanks to its anti-shear properties. They are also used in healthcare as a basic ingredient for consumer products that regulate intestinal gas production or as a substitute for the vitreous in retinal removal surgeries. PDMS is also a popular test material for new rheological theories, and for novel measuring methods and devices. Some rheometer manufacturers (e.g., Anton Paar) use PDMS also as calibration liquids [11].

2 Rheology of Fluids

Rheology is partially related to chemical engineering and partially to solid state physics and aiming to better understand the flow and deformation of liquids and soft solids due to effect of stresses in the material generated by external forces.

The science of rheology is capable of understanding and quantifying the basic nature of the matter, as well as classifying materials with similar behavior and flow-deformation properties. The fundamental relation of rheology, also known as the constitutive or material equation is written in Eq. (1),

$$\tau = \eta \cdot \dot{\gamma} \tag{1}$$

where τ: shear stress [Pa], $\dot{\gamma}$: shear rate [1/s], the velocity gradient associated with the displacement between the adjacent fluid layers, and η: dynamic viscosity [Pas], the resistance of the fluid to deformation, the degree of internal friction between the adjacent fluid layers.

The actual value of viscosity mostly depends on temperature, pressure, shear rate, and time. The most important amongst them is the shear rate in categorization point of view. If the shear stress arising in the fluid depends only on the shear rate at the given time, then one can talk about a simple fluid, while in other cases it is called a viscoelastic fluid. Silicone oil is a viscoelastic and pseudoplastic fluid or shear-thinning fluid where the shear rate increase results in a viscosity decrease according to a given function [12]. Even though silicone oils are viscoelastic fluids [12] and their complex non-Newtonian behavior is usually described both with viscous and elastic terms, considering the filling process of visco-dampers as a relative slow fluid dynamic problem involving small cavities and narrow channels (where the flow times are measured in minutes at room temperature), the authors found the role of elastic terms negligible in this work and focused only on the description of viscous behavior.

1.3 Rheological Measurement Methods

In order to facilitate, and to perform the rheological measurements for determining the basic rheological quantities (shear rate, shear stress, viscosity) and to make the results more reliable and easier to evaluate, simple geometric conditions are chosen for the measurement. Two basic types of rheological measuring devices are used in practice [13]:

- Capillary viscometers – flow in a channel with a simple (usually circular) cross-section
- Rotational viscometers – flow between parallel plates (PP geometry), flow between cone and plate (CP geometry) or flow between concentric cylinders (CC geometry)

The great advantage of rotational viscometer is that it can be used to determine the forces acting perpendicular to the flow direction, i.e., the normal stresses. Based on these, the flexibility of the fluid can be quantitatively characterized.

In case of PP geometry, the sample is located between two parallel plates at an adjustable distance from each other. The bottom plate is fixed while the top plate rotates. During measurement, a steady, laminar and isothermal flow is formed, however, the shear rate in the fluid is not constant. The advantage of this arrangement is the simplicity of sample loading, the possibility to easily adjust the size of the gap between the plates and the possibility to measure shear stresses in the normal direction.

In order to develop a silicone oil material model suitable for calculating the filling process of visco-dampers, rheological analyses have to be carried out, which include temperature- and shear rate-dependent viscosity measurements of a pseudoplastic, high-viscosity liquid. The purpose of the measurements is to determine the basic rheological quantities of silicone oil in a known and prescribed temperature and shear range. Instead of describing the complex viscoelastic behavior of the oil, it is enough only to model the viscous behavior of the silicon oil for the filling process analysis. This requires a rotational viscometer and is suitable for performing constant shear test and simple shear test. Constant shear rate test means that one part of the equipment rotates at a constant, relatively low speed. This method is used to determine the basic rheological quantities, such as shear rate and shear stress and the viscosity can be calculated from them. In case of simple shear test, the shear rate or the shear stress is increased while the other quantity is measured. During the filling process of a TVD, the silicone oil is exposed to static and short-term (few minutes long) stress, so it is unnecessary to carry out oscillation and creep tests. Kőkuti's work [12] reveals the fact, that the necessary time required for this type of silicone oil to test its thixotropy can be measured in hours, so the time dependence analysis of silicone oil's viscosity is out of scope in this work.

1.4 Rheological Measurement Results of AK 1 000 000 STAB Silicone Oil

The rheological measurement results of AK 1 000 000 STAB silicone oil, presented in the current section, are valid at atmospheric pressure. The reason behind the lack of pressure-dependent viscosity analysis is that it has negligible effect, the recently investigated filling process operates at ambient pressure and the pressure-viscosity diagram available in the silicone oil's product brochure [14] that provides enough information about the pressure dependence of the viscosity. According to this diagram, all types of silicone oil react in the same way to pressure changes. Even though the viscosity of the silicone oil increases as pressure increases, silicone oils are much less sensitive to pressure changes than mineral oils. In fact, their viscosity experienced under ambient conditions (10^5 Pa) is doubled at 450×10^5 Pa [14].

The rheological measurements of the AK 1 000 000 STAB silicone oil, chosen as the subject of the study, were carried out by an Anton Paar Physica MCR 302 rotational rheometer. In order to minimize the loss of friction and resulting inaccuracy during the rotation of the measuring head, the instrument is equipped with a fine porous carbon coated air bearing; with an active temperature-controlled, electronically commutated permanent magnet synchronous motor (EC-motor), with an integrated normal force sensor and with a high-resolution optical encoder. This modular compact rheometer (MCR) is a "shear stress-controlled rheometer", which means, that the rheometer calculates the shear stress applied to the fluid sample from the instantaneous value of the electric motor's current that moves the rotating part of the measuring geometry, and at the same time, this current is controlled in such a way, that the prescribed time-dependent shearing is realized in the fluid sample based on the data of the optical angular rotation sensor mounted on the same axis as the motor. The torque range that can be applied with the instrument is 0.5 nNm-230 mNm; which has an angular velocity range of 0-314 rad/s (3000 RPM) and the target temperature range of the fluid sample is -40°C to 200°C.

The main components of the measuring system are displayed in Figure 2:

- MCR 302 rheometer (1) with cooling/heating air volume flow rate regulator, air- and fluid circulated Peltier system (2) and PP25 measuring head geometry (3);
- Compressor (4) to keep the volume flow rate of the cooling/heating air guided to the air-circulated Peltier element at a constant value;
- Thermostat and heat exchanger (5) filled with hydraulic fluid to control the temperature of cooling/heating air and cooling/heating fluid;
- Hydraulic pump (6) for circulating the cooling/heating fluid in the measuring system;
- Desktop computer (7) for controlling the measuring instrument via RheoCompass software;
- Additional components (8) such as calibration fluid, sample dispenser, sample handling and removal spoons, acetone cleaning agent, wipes.

Based on a private communication with the Anton Paar Support Team, in case of high-viscosity samples (e.g., silicone oil with an initial viscosity of around 1000 Pas), measurements above the shear rate of 1000 1/s are laden with greater error than measurements below this threshold value. Furthermore, in the high shear rate range (1000 1/s already falls within this range), simple shear tests are proved to be more accurate than oscillation tests. The accuracy of the instrument is claimed to be 1.5% for CC geometry and 3% for PP and CP geometries.

The theoretical considerations written in subsection 2.1 and 2.2 were taken into considerations as well as the properties and limitations of the available measuring instrument. According to this background, the authors prepared and implemented the following measurement plan.

Figure 2
Measuring system with the main components used in the rheological measurements

The goal of the rheological analysis is to measure the dynamic viscosity curves of the AK 1 000 000 STAB silicone oil with the Anton Paar Physica MCR 302 rotational rheometer at 100 points in the shear rate range of 0 1/s and 1000 1/s; between -40°C and 200°C at 14 different temperatures; at atmospheric pressure; applying a PP25 measuring head geometry (according to the standard protocols, one has to work with a fluid sample amount of 0.5 ml for each measurement).

Main steps of the measurement plan:

I. The first step of the rheological analysis is to determine the accuracy and precision of the measuring instrument with a given measurement gap, and to determine the minimum required number of measurements for a given measurement case (in other words: how many times the viscosity measurement must be repeated at a given temperature and shear rate) on a suitably chosen confidence level.

II. The second step of the rheological analysis is to determine the time required for the entire silicone oil sample placed on the measuring plate to reach the desired target temperature through cooling or heating. The initial viscosity at the prescribed 14 different test temperatures is also recorded.

III. The third step of the rheological analysis is to carry out the dynamic viscosity measurements in the shear rate range of 0 1/s and 1000 1/s at all 14 different temperatures repeated the required number of times.

Details of the measurement steps:

In the first step, constant shear rate tests were repeated 30 times at a shear rate of 0.1 1/s, at 25°C and with a gap size of h = 0.2 mm. This gap size provided the most accurate results for the initial dynamic viscosity of the samples at 25°C according

to the product brochure [14]. The absolute value of the deviation of the arithmetic mean of the 30 measured and recorded viscosity values from the exact (catalogue) value gives the accuracy of the measuring instrument (A = 39 Pas). The corrected empirical standard deviation of the measured values from the mean indicates the precision of the measuring instrument (P = 12.65 Pas). If the confidence level is chosen to be 95% (this means that in 95% of the performed measurements the measured result must fall within the accuracy range of the measuring instrument, i.e., within the error limit of the instrument), the confidence factor (k) can be found from the standard normal distribution table for the 95% confidence level (in this case k = 1.96). The minimum required number of measurements (n) is given by Eq. (2) based on the measurement techniques [15].

$$n = \left(k \cdot \frac{P}{A}\right)^2 = \left(1.96 \cdot \frac{12.65\ Pas}{39\ Pas}\right)^2 = 0.404 \tag{2}$$

Rounding up n = 1 is gained, therefore it is theoretically sufficient to measure only once in each measurement case with the high-precision measuring instrument presented at the beginning of the subsection. Regardless of the gained number, each measurement was performed 5 times based on 'one measurement is not a measurement' principle.

In the second step, after different cooling/warming waiting times, constant shear rate measurements were carried out at a shear rate of 0.1 1/s with a gap size of h = 0.2 mm for 5 minutes at each investigated temperature. The viscosity-shear rate diagrams were selected, on which the viscosity of the sample remained almost constant during the 5 minutes of the measurement. During the measurements, the farther the measurement temperature was from 25°C room temperature, more time was needed for all regions of the sample to reach the target temperature and for the initial viscosity value to remain unchanged during the measurement. Finally, 15 minutes is proved to be the necessary minimum waiting time, which ensures that all regions of the sample will take the target temperature in all investigated temperature cases. The initial viscosity values measured at each investigated temperature are listed in Table 1.

In the third step, simple shear tests were performed at all 14 investigated temperatures with a gap size of h = 0.2 mm, with a cooling or heating waiting time of 15 minutes, repeated 5 times, in the shear rate range of 0 1/s and 1000 1/s (logarithmically increasing value distribution in $N = 100$ points) and with a shearing time interval between 10 s and 0.01 s (with a logarithmically decreasing value distribution). In this step, a total of 7000 pieces of measured viscosity data were stored and evaluated. The following quantities were recorded: Elapsed time of measurement [s], Temperature [°C], Dynamic viscosity [Pas], Shear rate [1/s], Shear stress [Pa], Shear strain [%], Torque [Nm], and Normal force [N].

At each investigated temperature, the shear stress and dynamic viscosity values recorded at the same shear rate (measured 5 times) were averaged. Figure 3 depicts

the measured points of the rheological curves obtained by averaging the results and present well the non-Newtonian behavior of the analyzed silicone oil.

Table 1
Measured initial dynamic viscosity values of AK 1 000 000 STAB silicone oil

Temperature [°C]	Dynamic viscosity [Pas] at 0.1 1/s shear rate	Temperature [°C]	Dynamic viscosity [Pas] at 0.1 1/s shear rate
-40	5968.995	80	459.890
-20	2842.767	100	377.767
0	1652.900	120	302.737
20	1102.225	140	244.270
25	1005.600	160	203.750
40	786.670	180	178.753
60	590.630	200	131.145

The rheological measurement results clearly reveal the pseudoplastic nature of the silicone oil. The shear stress generated in the fluid increases with the increasing shear rate, but its actual dynamic viscosity decreases (see Figure 3). The measurement results also confirm the fact that silicone oil is a thermorheological simple fluid (fulfils the Time Temperature Superposition rule) since its rheological characteristics behave in the same way at all temperatures, the rheological characteristic curves are only shifted in the vertical direction based on the temperature.

1.5 Commonly Used Viscosity Models for Pseudoplastic Fluids

Based on the literature, eight most used pseudoplastic viscosity models were selected for comparison. These are the Carreau, Carreau-Yasuda, Cross, Johnson, Meter, Münstedt, Powell-Eyring and Power-Law [16]. The formula of each investigated viscosity model is presented by Eqs. (3)-(10) in Table 2. Nonlinear regressions were applied for each model separately, in order to identify the unknown parameters. Table 4 compares the fitting accuracy of each model to the measurement data based on several aspects. A viscosity model is then highlighted by the authors that is found to be the most appropriate for CFD calculation of the visco-damper's filling process based on evaluating aspects.

The formulas of the viscosity models are considered as a fitting function and the averaged measured dynamic viscosity values presented in Figure 3 are used to perform the model parameter identification by applying least squares method based nonlinear quasi-Newtonian regression tracing back the solution into an extremum search in iterative way. The functional minimum is gained by making the partial derivatives of the functional equal to zero.

Table 2
Investigated non-Newtonian viscosity models

Model	Formula	Eq.
Carreau	$\eta(T,\dot{\gamma}) = \eta_\infty(T) + \frac{\eta_0(T) - \eta_\infty(T)}{[1 + (\lambda(T)\cdot\dot{\gamma})^2]^{\frac{1-n(T)}{2}}}$	(3)
Carreau-Yasuda	$\eta(T,\dot{\gamma}) = \eta_\infty(T) + \frac{\eta_0(T) - \eta_\infty(T)}{[1 + (\lambda(T)\cdot\dot{\gamma})^{a(T)}]^{\frac{1-n(T)}{a(T)}}}$	(4)
Cross	$\eta(T,\dot{\gamma}) = \eta_\infty(T) + \frac{\eta_0(T) - \eta_\infty(T)}{1 + (\lambda(T)\cdot\dot{\gamma})^{1-n(T)}}$	(5)
Johnson	$\eta(T,\dot{\gamma}) = \eta_\infty(T)\cdot e^{\left(\frac{1}{\frac{1}{ln\eta_0(T) - ln\eta_\infty(T)} + \lambda(T)\cdot\dot{\gamma}}\right)}$	(6)
Meter	$\eta(T,\dot{\gamma}) = \eta_\infty(T) + \frac{\eta_0(T) - \eta_\infty(T)}{1 + \left(\frac{\dot{\gamma}}{\dot{\gamma}_{1/2}(T)}\right)^{1-n(T)}}$	(7)
Münstedt	$log\eta(T,\dot{\gamma}) = A(T) + B(T)\cdot log\dot{\gamma} + C(T)\cdot(log\dot{\gamma})^2 + D(T)\cdot(log\dot{\gamma})^3 + E(T)\cdot(log\dot{\gamma})^4$	(8)
Powell-Eyring	$\eta(T,\dot{\gamma}) = \eta_\infty(T) + \left(\eta_0(T) - \eta_\infty(T)\right)\cdot\left(\frac{sinh^{-1}(\lambda(T)\cdot\dot{\gamma})}{\lambda(T)\cdot\dot{\gamma}}\right)$	(9)
Power-Law	$\eta(T,\dot{\gamma}) = K(T)\cdot\dot{\gamma}^{n(T)-1}$	(10)

The model parameters $P(T)$ in each viscosity model formula depends on the temperature, thus a model parameter can be approximated by an eight-order polynomial.

The correct value of the parameter coefficients c_i in each polynomial for each viscosity model is gained by polynomial regression. To enhance the convergence during the iteration, the unknown model parameters are set to be only non-negative values. As a result, in case of the Carreau-Yasuda model, the polynomials for the temperature dependent model parameters are shown in the Table 3, which satisfied the sign constraint and provided the least squares deviation in the measured shear rate range.

Based on Table 4, from the highest absolute difference's point of view, the Carreau-Yasuda viscosity model is found to be the most reliable. In terms of highest relative difference, the Cross model works better. Considering the average of relative differences, the Münstedt model provides smaller differences to the measured viscosity values.

It is also important to investigate how does each viscosity model behave above the measured shear rate region (above 1000 1/s). If the calculated viscosity values diverge to infinite or fluctuate, they can cause numerical error during the CFD specific limiters.

Table 3
Identified model parameter polynomials for the selected viscosity model

Model	Polynomials of the temperature-dependent model parameters
Carreau-Yasuda	$\eta_\infty(T) \cong 0$ $\eta_0(T) = 3.55208E\text{-}14 \cdot T^8 - 3.04511E\text{-}11 \cdot T^7 + 1.06686E\text{-}08 \cdot T^6 - 1.98312E\text{-}06 \cdot T^5 + 2.15475E\text{-}04 \cdot T^4 - 1.48606E\text{-}02 \cdot T^3 + 0.77383 \cdot T^2 - 38.54805 \cdot T + 1662.14402$ $\lambda(T) = 2.97570E\text{-}19 \cdot T^8 - 3.44028E\text{-}16 \cdot T^7 + 1.49141E\text{-}13 \cdot T^6 - 3.26433E\text{-}11 \cdot T^5 + 4.06031E\text{-}09 \cdot T^4 - 3.17045E\text{-}07 \cdot T^3 + 1.84093E\text{-}05 \cdot T^2 - 9.51978E\text{-}04 \cdot T + 3.99458E\text{-}02$ $a(T) = 1.20867E\text{-}08 \cdot T^3 - 5.12543E\text{-}06 \cdot T^2 + 7.77623E\text{-}04 \cdot T + 0.84655$ $n(T) = 5.18491E\text{-}06 \cdot T^2 - 1.19737E\text{-}04 \cdot T - 6.35822E\text{-}03$

Because of this fact, Münstedt model is excluded from the selection as the model would result divergence and numerical error at higher shear rates. As far as Cross model is concerned, not only the highest absolute difference but also the average of relative differences is higher compared to Carreau-Yasuda model, thus Carreau-Yasuda viscosity model is selected for the most reliable and accurate viscosity model for the CFD calculations that can be implemented into ANSYS FLUENT environment without limiters in form of a user-defined function.

Figure 3 shows not only the averaged measured dynamic viscosity values (dots) but also the dynamic viscosity curves (lines) calculated by the Carreau-Yasuda model for AK 1 000 000 STAB silicone oil at each measured temperature above the investigated shear rate range.

Table 4
Comparison of the investigated viscosity models based on different aspects

Model	**Highest absolute difference to the measurement**	**Highest relative difference to the measurement**	**Average of relative differences (deviation)**	**Model behavior at higher shear rates (above 1000 1/s)**
Carreau	815.659 Pas	15.702%	4.913% (3.365%)	viscosity monotonically decreases to zero
Carreau-Yasuda	62.483 Pas	10.757%	1.105% (1.381%)	viscosity monotonically decreases to zero
Cross	195.9 Pas	6.399%	1.167% (1.033%)	viscosity monotonically decreases to zero
Johnson	1110.325 Pas	34.207%	4.137% (4.603%)	viscosity converges to a value between 18 and 55 Pas
Meter	125.326 Pas	7%	1.383% (1.101%)	viscosity monotonically decreases to zero

Münstedt	179.377 Pas	7.723%	1.082% (1.079%)	viscosity diverges to ± ∞
Powell-Eyring	357.276 Pas	32.66%	5.436% (5.017%)	viscosity monotonically decreases to zero
Power-Law	2818.624 Pas	88.675%	10.962% (15.928%)	viscosity monotonically decreases to zero

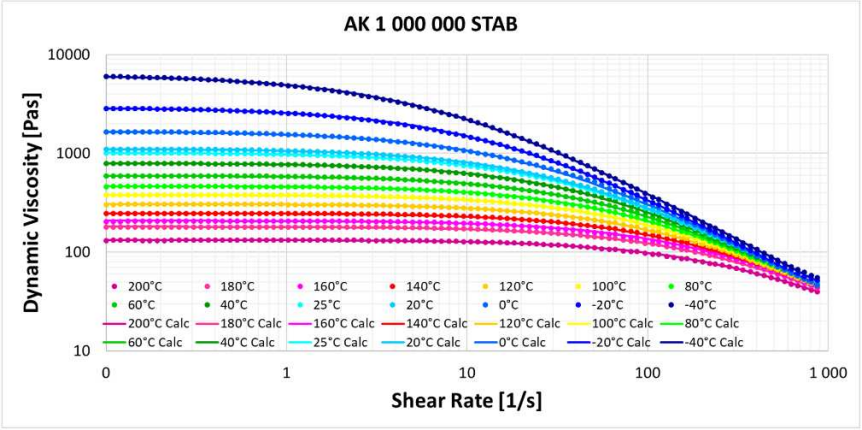


Figure 3
Measured (dots) and Carreau-Yasuda calculated (lines) dynamic viscosity values

3 Computational Fluid Dynamic Analysis of the Filling Process of a Visco-Damper

In design and development phases of TVDs the estimation and calculation of fluid flow, the thermal control of silicone oil, the reduction of filling time and the management of silicone oil's degradation level are important aspects for increasing the lifetime of the product and increasing the revenue. Hence, a simplified, small-sized, narrow damper gap geometry has been used to investigate the filling process at different discharge pressures.

The filling is carried out on a special filling bench, which is capable of creating a near-vacuum condition in the damper gap (around 100 Pa remaining absolute pressure in the gap) as a first step, and then injecting a pre-programmed amount of oil into the damper gap on a constant high pressure as a second step. As the third and last step, the plug is welded into the filling hole. The oil in the visco-damper is allowed to undergo minimal degradation from thermal point of view to provide maximum service life under operating conditions. For this reason, it is necessary to know the maximum oil temperature value recommended by the oil manufacturer, at which the oil still can be used safely under operation.

AK 1 000 000 STAB thermally stabilized silicone oil is used for high-performance visco-dampers manufactured by Wacker Chemie AG which is a German company headquartered in Munich and is the market leader among many manufacturers producing silicone oils. According to the manufacturer's technical data sheet [17], AK 1 000 000 STAB silicone oil can be used at operating temperatures higher than 150°C. According to the Wacker Chemie AG product brochure of the AK series silicone [14], this type of oils shows excellent long-term heat resistance up to 150°C in the presence of air. Above 150°C, however, permanent viscosity change occurs due to the oxidation in the fluid.

Based on the items above, the task is to examine the filling process on the damper gap geometry of a small-sized visco-damper by using a reliable and accurate temperature- and shear rate- dependent viscosity model developed in subsection 2.3 and using a temperature-dependent density model found in the product brochure [14] with specific heat, thermal conductivity and surface tension. The filling process is calculated at four different filling inlet overpressures (10^5 Pa, 5×10^5 Pa, 10×10^5 Pa and 20×10^5 Pa) and the maximum allowed filling inlet temperature of the oil is determined for all pressure cases, so that during the filling process the oil's temperature will not exceed in any point of the fluid the 150°C and the minimal degradation of the oil is ensured.

The authors were intended to perform 'works-case analysis' to gain highest possible oil temperatures from the filling process. Because of this fact, in each investigated filling case the heat transfer on the walls must be considered adiabatic by allowing no cooling effect on the damper housing.

The above-mentioned analyses were completed by CFD calculations based on the finite volume method. 3D, transient, multiphase, coupled flow and thermal simulation model is developed in the ANSYS FLUENT software environment to analyze the filling process of the damper.

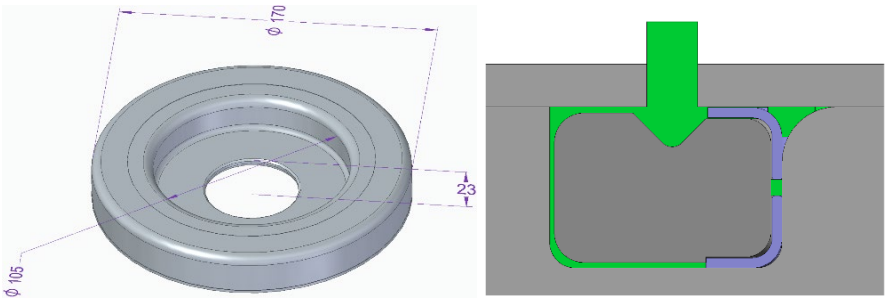


Figure 4
Main dimensions (left) and the cross-section of the investigated visco-damper with the damper gap highlighted (right)

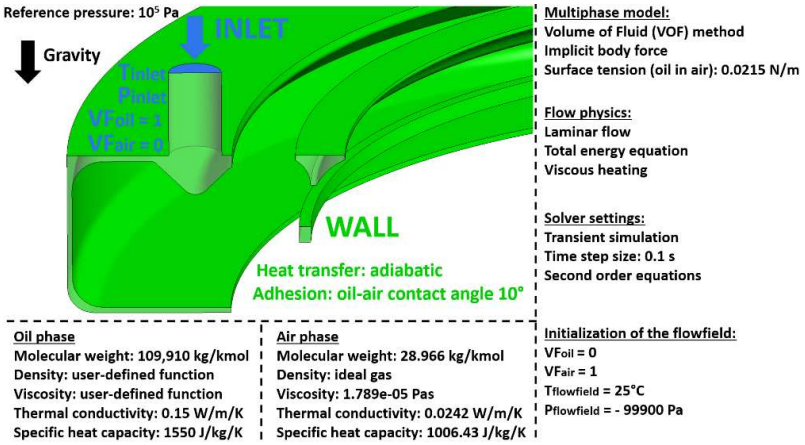


Figure 5

CFD simulation model with main boundary conditions and applied settings

1.6 Geometry and the Boundary Conditions for the Filling Simulation

A small-sized visco-damper has been selected for the CFD analysis that is shown on the left side of Figure 4 with its main dimensions in mm. The silicone oil's flow field (damper gap) is highlighted by green on the right side. The CFD simulation setup of the damper's filling process with the boundary conditions and applied models are depicted in Figure 5. In each case, the filling process is performed as long as 25 g amount of silicone oil is filled into the damper gap domain.

1.7 Numerical Results and Discussion

The main outcomes of the numerical calculations for each investigated filling case are collected in Table 5, and a qualitative result about the test case at 20×10^5 Pa inlet pressure and at 102°C inlet temperature is shown in Figure 7. The highest allowable oil inlet temperature values for each investigated filling case are highlighted by the pressure-case's color in 'Oil inlet temperature' row of Table 5.

There is a temperature increase in the silicone oil during the filling process in each investigated simulation case. The reason behind this temperature development is explained as follows. On the one hand, while the oil is filled into the narrow damper gap on a relative high filling pressure, internal friction occurs among the fluid layers. This mechanical energy is converted into heat and increases the oil's total temperature. On the other hand, as the oil spreads in the gap the volume of the gap-air reduces and its temperature increases (suffers from compression).

Table 5
CFD simulation results of the investigated filling cases

Oil inlet overpressure [Pa]	10^5		5×10^5		10×10^5		20×10^5	
Oil inlet temperature [°C]	25	**119.4**	25	**107.7**	25	**104**	25	**102**
Oil peak temperature [°C]	25.4	150	29.5	150	54.2	150	86.6	150
Maximal oil velocity [m/s]	0.06	0.12	0.29	0.6	0.95	1.65	7.35	9.8
Filled amount of oil [g]	25	25	25	25	25	25	25	25
Filling time required [s]	325	124.9	96.4	41.9	44.2	21.6	23.2	18.4

A part of the additional thermal energy originated from the compression of the gap-air transfers into the silicone oil through the oil-air interfaces.

The filling process is found to be the slowest in case the inlet overpressure is selected for 10^5 Pa and the oil inlet temperature is set to room temperature (25°C). In this case, 325 s is needed to fill 25 g oil into the gap and the silicone oil suffers only 0.4°C total temperature increases during the filling process. If the inlet overpressure is set to 20×10^5 Pa and the filling temperature is set to 102°C, the shortest filling time (18.4 s) can be reached for the same amount of silicone oil to be filled in such a way that the oil's total temperature does not exceed the maximal allowed 150°C value in any point of the oil phase domain (oil's total temperature increase is 48°C) during the filling process.

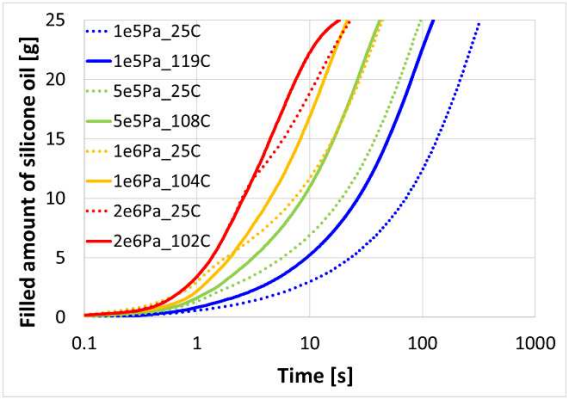


Figure 6
Filling mass characteristics

Figure 6 displays the filled amount of silicone oil's mass increase over time considered as the filling mass characteristics of the visco-damper (related to AK 1 000 000 STAB silicone oil and 25 g target mass). Figure 7 depicts the calculated filling process for 20×10^5 Pa oil inlet overpressure and 102°C oil inlet temperature simulation case in two different timesteps by coloring the oil phase domain according to the oil's total temperature distribution. The t = 3.5 s means the moment when exactly the half of the 25 g amount of oil is entered into the damper gap while t = 18.4 s is the last moment of the filling process when exactly 25 g amount of oil is pushed into the damper gap. As the time passes during the filling process, the pressure difference between the gap-air and the inlet decreases while the high-pressure zone is spreading in the oil and the oil slows down gradually, it is exposed to lesser shear, thus the dynamic viscosity of the oil shows a gradual increase to its maximum (initial) value (in case of 102°C this initial value is close to 380 Pas based on Table 1). At the end of the filling process, the majority of the oil phase domain reaches the minimal velocity (below 0.1 m/s) and its dynamic viscosity is increased back to the initial viscosity value. The oil temperature increase can be observed mainly along the oil-air interfaces (additional heat transferred from the compressed gap-air) and along the walls of the inertia-ring which region is well insulated by the oil cover.

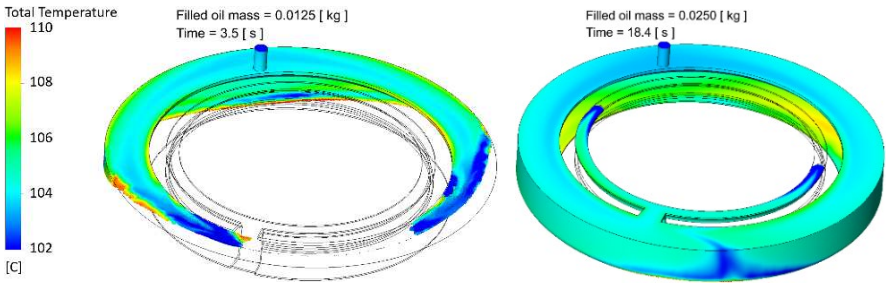


Figure 7
Total temperature distribution in the silicone oil at the inlet total pressure 20×10^5 Pa, at the inlet total temperature 102°C, and at t = 3.5 s (left) and at t = 18.4 s (right)

Conclusions

Dynamic viscosity measurements have been carried out on AK 1 000 000 STAB silicone oil samples, in the form of simple shear tests, using a high-precision rotational rheometer, to quantify the temperature and shear rate dependence of the mentioned silicone oil's non-Newtonian viscous behavior. Eight commonly used pseudoplastic non-Newtonian viscosity model have been selected for model parameter estimation and comparison to develop a reliable, accurate and easy-to-implement viscosity model valid in -40°C to 200°C temperature and 0 1/s to 1000 1/s shear rate ranges, for CFD calculations to analyze the filling process of an existing, small-sized visco-damper under eight different filling conditions. Carreau-Yasuda viscosity model is found to be the most appropriate for describing the viscous behavior of the analyzed high-viscosity silicone oil in CFD simulations.

The outcome of the numerical analysis shows that in case the silicone oil is filled into the damper gap, at highest allowed temperature (instead of 25°C), the filling time can be shortened:

- at 10^5 Pa inlet overpressure by 61.6%
- at 5×10^5 Pa inlet overpressure by 56.5% (by 87.1% related to 10^5 Pa – 25°C)
- at 10×10^5 Pa inlet overpressure by 51.1% (by 93.4% related to 10^5 Pa – 25°C)
- at 20×10^5 Pa inlet overpressure by 20.7% (by 94.3% related to 10^5 Pa – 25°C)

In the above presented cases, the oil's total temperature increase remains under the allowed 150°C and no permanent thermal degradation of the silicone oil occurs. The next step of this work is to introduce filling process calculation method, by neutron radiography measurements, to investigate the impact of different damper assembly configurations (changing the position of bearing gap and the orientation of the inertia-ring) on the filling time, drawing more filling mass characteristics and filling velocity characteristics, for different assembly configurations and to optimize the filling process, for additional filling time reductions, with minimal oil degradation.

Acknowledgement

This work was supported by the Pro Progressio Foundation. The authors would like to express their special thanks to Bálint Heizer sales specialist and to the Anton Paar Hungary Kft. for providing them the Anton Paar MCR 302 rotational rheometer and for enabling them the possibility to carry out the presented rheological measurements.

References

[1] Zhu, X., Jing, X., Cheng, L.: Magnetorheological fluid dampers: A review on structure design and analysis, Journal of Intelligent Material Systems and Structures, 2012, 23(8), pp. 839-873

[2] Absolute Reports: Global Torsional Vibration Damper Market 2019 by Manufacturers, Regions, Type and Application, Forecast to 2024, 2019, Pune, India, URL: https://www.absolutereports.com/global-torsional-vibration-damper-market-13851137 (Accessed: 21.11.2022)

[3] Wojciech Homik: Torsional vibration silencers used in vessels propulsion systems, Scientific Journals of the Maritime University of Szczecin, 2014, 40(112), pp. 9-16

[4] Hasse&Wrede GmbH: Visco Damper After Sales Service, serviceflyer, Berlin, Germany, 2013, URL: https://www.hassewrede.com/media/documents/Serviceflyer.pdf (Accessed: 02.10.2022)

[5] Vaclav Pistek, Lubomir Klimes, Tomas Mauder, Pavel Kucera: Optimal design of structure in rheological models: an automotive application to

dampers with high viscosity silicone fluids, Journal of Vibroengineering, 19(6), 2017, pp. 4459-4470

[6] W. Homik, A. Mazurkow, P. Woś: Application of a Thermo-Hydrodynamic Model of a Viscous Torsional Vibration Damper to Determining Its Operating Temperature in a Steady State, Materials, 2021, 14(18), 5234

[7] Nicholas J. Fendinger: Polydimethylsiloxane (PDMS): Environmental Fate and Effects, Organosilicon Chemistry Set: From Molecules to Materials, WILEY-VCH Verlag GmbH & Co. KGaA, 2005, pp. 626-638, DOI: https://doi.org/10.1002/9783527620777.ch103c

[8] H. Miyahara, A. Nakajima, J. Wada and S. Yanabu: Breakdown Characteristics of Combined Insulation in Silicone Oil for Electric Power Apparatus, 2006 IEEE 8th International Conference on Properties & applications of Dielectric Materials, 2006, pp. 661-664

[9] J. A. Brydson: 29 - Silicones and Other Heat-resisting Polymers, Plastics Materials (Seventh Edition), Butterworth-Heinemann, 1999, pp. 814-852

[10] Christian Stenzel: Deployment of precise and robust sensors on board ISS—for scientific experiments and for operation of the station, Analytical and Bioanalytical Chemistry, 2016, 408, pp. 6517-6536

[11] Brandstaetter M.: MCR Series: Correct Adjustment of the Rheometer and Measurement of Standard Samples using RheoPlus, Anton Paar Application Report C92IA005EN-A, 2013

[12] Z. Kőkuti, J. Kokavecz, A. Czirják, I. Holczer, A. Danyi, Z. Gábor, G. Szabó, N. Pézsa, P. Ailer, L. Palkovics: Nonlinear Viscoelasticity and Thixotropy of a Silicone Fluid, Annals of Faculty Engineering Hunedoara - International Journal of Engineering, 2011, 9(2), pp. 177-180. URL: https://annals.fih.upt.ro/pdf-full/2011/ANNALS-2011-2-35.pdf (accessed: 02.10.2022)

[13] Béla Pukánszky: Rheology, laboratory notes, Budapest University of Technology and Economics, Faculty of Chemical Technology and Biotechnology, Laboratory of Plastic and Rubber Industry, 2019, pp. 43-66. URL: https://mua.bme.hu/hallgatok/letoltesek/NYILVANOS_TARTALOM/laborleiratok/Anyagtudomany_BSc/5.%20Reol%c3%b3gia.pdf (accessed: 02.10.2022)

[14] Wacker-Chemie GmbH: Wacker Silicone Fluids AK, product brochure, Wacker-Chemie GmbH Silicones Division, 2002, URL: https://www.behlke.com/pdf/wacker_silicone_oil.pdf (accessed: 02.10.2022)

[15] Károly Havancsák: Management and Evaluation of Measurement Data, lecture notes, Eötvös Loránd University, Faculty of Science, Department of Materials Physics, Typotex, 2012, pp. 148-153, URL: http://etananyag.ttk.elte.hu/FiLeS/downloads/EJ-Havancsak_Meresi-adatok.pdf (accessed: 21.11.2022)

[16] Göttfert Werkstoff-Prüfmaschinen GmbH: Modeling of Viscosity Data, 2022, URL: https://www.goettfert.com/application-knowledge/rheo-info/for-capillary-rheometer/modeling-of-viscosity-data (accessed: 07.12.2022)

[17] Wacker-Chemie GmbH: WACKER® AK 1 000 000 STAB, URL: https://www.wacker.com/h/en-us/silicone-fluids-emulsions/linear-silicone-fluids/wacker-ak-1-000-000-stab/p/000016048 (accessed: 02.10.2022)

Cybersecurity Attack Detection Model, Using Machine Learning Techniques

İsa Avcı[1] and Murat Koca[2]

[1]Department of Computer Engineering, Faculty of Engineering, Karabuk University, Kılavuzlar Mahallesi 413. Sokak No: 7, 78000, Merkez, Karabuk, Turkey, isaavci@karabuk.edu.tr

[2]Department of Computer Engineering, Faculty of Engineering, Van Yuzuncu Yil University, Kampüs, 65080, Tuşba, Van, Turkey, muratkoca@yyu.edu.tr

Abstract: Millions of people use the web every day, in this age of technology and the internet. Protecting the privacy and security of these users is a significant challenge for cybersecurity developers. With tremendous technological advancements, there is a noticeable improvement in the cyber-attackers' capabilities. At the same time, traditional Intrusion Detection Systems (IDS) are no longer effective at detecting intrusions. After the tremendous competences achieved by Artificial Intelligence (AI) techniques in all fields, great interest has developed in its use in the field of cybersecurity. There have been many studies that use Machine Learning (ML)-based intrusion detection systems. Despite the strong performance of ML techniques in detecting malicious activities, some challenges still reduce accuracy of performance. Knowing the proper technique, as well as knowing the features, is essential for effective intrusion detection. Therefore, this study proposes an effective network intrusion detection system based on ML and feature selection techniques. The performance of four ML techniques, the Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and the Decision Tree (DT) systems for intrusion detection are explored. In addition, feature selection techniques are employed for the selection of important features. Among the techniques used, the RF technique achieved the best performance, outperforming other techniques, with an accuracy of 99.72%. This study elaborates on the detection of malicious and benign cyber-attacks, with a new-level, high accuracy.

Keywords: cybersecurity; intrusion detection; DDoS attacks; machine learning; feature selection techniques

1 Introduction

All technological developments, including smartphones, computers, communication systems and IoT devices lead to the development of internet networks, throughout the world [1]. Recent studies estimate more than 5 billion

smart devices and 3 billion Internet users worldwide [2]. As a result of this great use of Internet networks, massive amounts of data are generated every second, which led to the creation of significant security challenges to protect data from the many challenges facing cybersecurity developers [3]. Cybersecurity protects computer systems and networks from unauthorized access [4]. Cybersecurity is a backbone for all types of companies, governments, and even people to secure data and maintain privacy. People send and receive data over the internet, which can be hacked and manipulated by strangers [5]. Cyber-attacks in 2017 caused damages of $5 billion and will only increase in the future; for example, damages were estimated to reach $6 trillion, annually, by 2021 [6].

Distributed Denial of Service (DDoS) attacks are among the most common cybersecurity attacks. DDoS is a lethal weapon that overwhelms a server or network by sending large amounts of packets, which floods the servers and causes the service to stop [7]. In recent years, DDoS attacks have witnessed an alarming increase. In February 2020, Amazon Web Services (AWS) customers suffered a severe crash when a DDoS attack targeted Amazon Simple Storage (S3) and other services, shutting them down for approximately 8 hours [8]. It is one of the most significant DDoS attacks, with a capacity of 2.3 terabytes per second. According to a Security Week article, researchers have discovered that the average number of separate DDoS attacks infecting the internet daily is about 28,700 [9]. Therefore, there was a need to create a reliable system for detecting cybersecurity attacks. Cybersecurity developers aim to create an effective IDS that can identify known and new attacks and threats with high accuracy and a low false alarm rate [10]. Although many methods are available for intrusion detection, the increase in the effectiveness of recent attacks and the evolution of attack methods, especially DDoS attacks, requires effective intelligent methods to detect them. Whereas, traditional IDSs are no longer effective in intrusion detection [11].

The use of artificial intelligence techniques in the field of cyber security has become mandatory with great success in every field. AI and ML techniques provide a tremendous ability to explore hidden models in big data, allowing them to help decision-making. ML techniques help detect and monitor attacks on network traffic activity. Many studies used different ML techniques to detect intrusion. However, there are still some shortcomings, including determining the appropriate ML technique for the intrusion detection process [12]. Basic and effective features have been selected to improve the performance of ML techniques in the detection of unauthorized inputs [13].

This study aimed to create an effective IDS based on machine learning and feature selection techniques. In our new approach, critical features that affect the classification result are selected as a basis for a more accurate classification process. The performance of four different machine learning techniques Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) are compared to find the appropriate technique for intrusion detection. One of the most effective ways to select features is to use a DT technique to determine the

feature's importance. The feature importance feature of the DT technique was found useful to determine the importance of each feature and its effect on the classification result. This proposed study is designed to detect malicious or benign traffic in DDoS cyber-attacks with a new high accuracy. In addition, the outliers in the data's normalization steps were standardized using regression methods. Data robustness and confidence intervals were also examined in detail.

The research article is organized as follows and related studies are discussed in Section 2. It describes the general methodology preprocessing steps and the application of machine learning techniques in Section 3. Section 4 deals with the results obtained and their discussion. Section 5 provides conclusions. Performance experiments are performed using the NSL-KDD dataset to determine whether the activity is innocuous [14]. This study evaluated the model's performance using the confusion matrix. The performances of the ML models were calculated in terms of accuracy, precision, sensitivity, specificity, and F1-Score.

2 Related Work

The significant development in technology and Internet of Things (IoT) devices leads to increasing use of Internet networks, which in turn requires effective protection methods to protect the privacy and security of the user. Artificial intelligence is one of the most promising approaches to countering cybersecurity threats. Many studies have used IDSs based on ML and Deep Learning (DL) techniques. This section discusses a set of studies that use ML and DL techniques to detect cybersecurity attacks.

Bindra and Sood explored 6 ML techniques LR, KNN, RF, NB, Linear SVM, and Linear Discriminant Analysis (LDA) to find out the best technique for detecting DDoS attacks [15]. ML techniques were tested on the CIC IDS dataset, where RF technology achieved the best performance with an accuracy of 96.5%, superior to the rest of the techniques. Also, Chavan et al. studied the performance of four ML techniques KNN, SVM, DT, and LR for detecting DDoS attacks [16]. Of all the techniques used, LR achieved the best accuracy with 90.4%, outperforming the rest. The ensemble method often produces a better accuracy rate than a base classifier. Therefore Das, Saikat, et al. proposed an ensemble model that combines 4 base machine learning ML techniques Multilayer Perceptron (MLP), SVM, KNN, and DT [17]. The performance experiments were tested on the NSL-KDD data set, where the ensemble classifier achieved better results than the individual classifiers used in the same study.

Kasim suggested using an Auto-Encoding (AE) method for selecting features and reducing dimensions to effectively classify traffic [18]. The AE is used to identify essential features, and the SVM classifier then detects a DDoS attack.

Performance experiments were performed on CICIDS2017 and NSL-KDD datasets, and the results showed the model's effectiveness for classifying traffic. Bhardwaj et al. introduced a method that combines well-stacked sparse AE for feature learning using a Deep Neural Network (DNN) to detect potential DDoS attacks [19]. The performance of AE and DNN was tuned by adjusting parameters to improve detection accuracy. Performance experiments were conducted on the CICIDS2017 and NSL-KDD datasets. The results showed that AE + DNN was superior to AE + SVM in the study with the NSL-KDD dataset, while the results were competitive, when using CICIDS 2017.

The high-efficiency DL techniques achieved in discovering big data have been many efforts to explore it in the field of cybersecurity. Al-Emadi et al. explored the performance of DL techniques in Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for network intrusion detection [20]. The performance experiments showed that the CNN technique is superior to the RNN technique when tested on the NSL-KDD dataset. Also, Abu Abu Al-Haija and Zein-Sabatto proposed CNN techniques for intrusion detection [21]. Performance experiments were conducted on the NSL_KDD dataset, whereby the CNN technique achieved a classification accuracy of 99.3% for intrusion detection. The performance of ML and DL techniques is highly dependent on data quality. Therefore, Xavier Larriva-Novo et al. explored various preprocessing techniques to classify traffic using the DNN technique [22]. This study showed, that by using preprocessing techniques, accuracy could be improved by up to 45%.

3 Proposed Methodology

This study proposes an effective network intrusion detection system based on ML and feature selection techniques. The performance measures are conducted four ML techniques RF, KNN, SVM, and DT for intrusion detection. In addition, feature selection techniques are used to identify essential features.

3.1. Research Design

Machine learning algorithms used in this part of this study are briefly summarized. In addition, in this section, performance criteria used in machine learning algorithms are given.

3.1.1 Decision Tree (DT)

The DT is a non-parametric supervised learning technique and one of the most influential classification techniques, which can be used for both classification and

regression problems. The decision tree structure is like the tree structure but from top to bottom, where the highest node in the tree represents the root. Each internal node represents a test on a feature, each branch indicates the result of the test, and each leaf node indicates a class label [23]. A Classification and Regression Tree (CART) is used to detect cyberattacks that generate binary trees and uses a Gini index function as a method for feature selection for classification problems in Equation (1).

$$Gini = 1 - \sum_{i=1}^{n} (p_i)^2 \tag{1}$$

3.1.2 Random Forest (RF)

RF is a supervised ML technique that can be used for both classification and regression. Since it grows many decision trees rather than a single decision tree in the model, RF is an ensemble learner. It means more trees which generates a more robust classifier. RF generates several CART, in which each tree is trained on a randomly selected subset of the original data set. The decisions of all the decision trees generated within the forest are aggregated, and a vote makes the classifying decision of most of the trees [24].

3.1.3 Support Vector Machine (SVM)

SVM is one of the most potent supervised ML models used for classification and regression problems, but it is commonly used in classification. The work of the SVM technique is to classify data by defining a hyperplane or a line separating two classes within a data set. To find the best line to separate the data, SVM calculates the distance between the points of the two different classes and determines the points closest to each hyperplane class, which are called support vectors, where the most significant margin separating the hyperplane and the support vectors are chosen [25].

3.1.4 K-Nearest Neighbors (KNN)

The KNN is one of the most straightforward ML techniques that can be used for both classification and regression problems. The KNN technique assumes that convergent objects are the same. In other words, similar things are close to each other. To classify a new condition KNN technique calculates the distance between the item to be classified and all the training data items. Then the best value of K is determined, which is the number of nearest neighbors of the element to be classified [26]. Usually, several values are tried to determine the optimal value of k. The majority vote of the neighbors determines the result of the classification.

To measure the distance between two points, the KNN technique used Euclidean distance in Equation (2).

$$Euclidean\ distance(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + \cdots + (x_{in} - x_{jn})^2} \quad (2)$$

3.1.5 Performance Measurements

The performance of ML techniques was evaluated using five quality measures that are Accuracy, Precision, Sensitivity, Specificity, and F1-Score. Malicious Samples are considered positive and represented by '1'. While benign samples are considered negative and represented by '0' [27]. All performance measure formulas are given in Equations (4-8).

- True Positives (TP): Malicious Samples have already been detected as malicious.
- True Negatives (TN): Benign Samples have already been detected as benign.
- False Positives (FP): Benign Samples have already been detected as malicious.
- False Negatives (FN): Malicious Samples have already been detected as benign.

$$\text{Accuracy} = \frac{(TN+TP)}{(TN+TP+FN+FP)} \quad (4)$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (5)$$

$$\text{Sensitivity} = \frac{TP}{(TP+FN)} \quad (6)$$

$$\text{Specificity} = \frac{TN}{(TN+FP)} \quad (7)$$

$$\text{F1 Score} = \frac{2*(\text{Precision} * \text{Sensitivity})}{(\text{Precision} + \text{Sensitivity})} \quad (8)$$

3.2. Data

In this study, the NSL-KDD dataset was used which is a clean and refined version of the University of New Brunswick KDD'99 dataset. A large amount of network traffic was collected to create the KDD dataset [28]. The NSL-KDD dataset consists of 42 features and 148,517 samples, these features are given in Table 1.

Table 1
Describes the features of the NSL-KDD data set

No	Features Names	No	Features Names	No	Features Names
1.	duration	**15.**	su attempted	**29.**	same_srv_rate
2.	protocol_type	**16.**	num_root	**30.**	diff_srv_rate
3.	service	**17.**	num_file_creations	**31.**	srv_diff_host_rate
4.	flag	**18.**	num_shells	**32.**	dst_host_count
5.	src_bytes	**19.**	num_access_files	**33.**	dst_host_srv_count
6.	dst_bytes	**20.**	num_outbound_cmds	**34.**	dst_host_same_srv_rate
7.	land	**21.**	is_host_login	**35.**	dst host_diff_srv_rate
8.	wrong fragment	**22.**	is_guest_login	**36.**	dst host same_src_port_rate
9.	urgent	**23.**	count	**37.**	dst_host_srv_diff_host_rate
10.	hot	**24.**	srv_count	**38.**	dst_host_serror_rate
11.	num_failed_logins	**25.**	serror_rate	**39.**	dst_host_srv_serror_rate
12.	logged_in	**26.**	srv_serror rate	**40.**	dst_host_rerror_rate
13.	num compromised	**27.**	rerror_rate	**41.**	dst host srv rerror rate
14.	root_shell	**28.**	srv_rerror rate	**42.**	class (malicious and benign)

The NSL-KDD dataset contains more than one type of cyber-attack, but this work focuses only on whether the traffic is malicious or benign in DDoS attacks. Fig. 2 shows the number of samples for the class type, malicious or benign. Briefly, this proposed study has been developed to detect malicious or benign traffic in DDoS attacks.

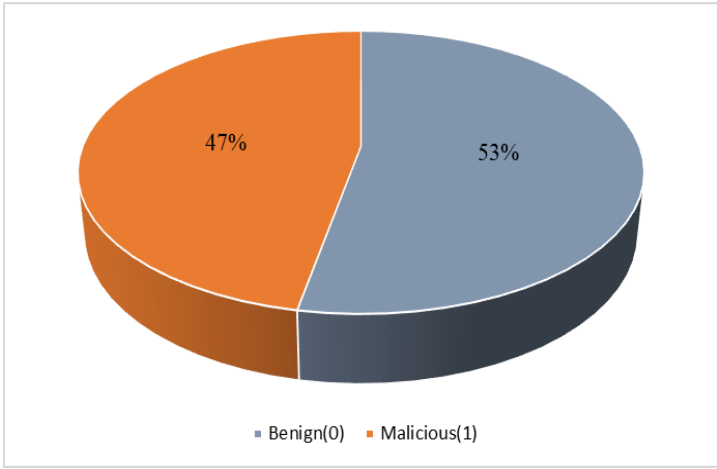


Figure 2
The percentage of each class

3.3. Proposed Methods

Data pre-processing is an important stage for ML techniques because the raw data often tends to be inconsistent and noisy and may contain missing, redundant, and irrelevant data. The efficiency of ML techniques depends mainly on the quality of the data provided [29]. So, to build a model with high performance and good accuracy, the pre-processing must be accurate. The pre-processing of NSL-KDD data in this study is summarized in the following steps. Figure 1 shows the proposed framework.

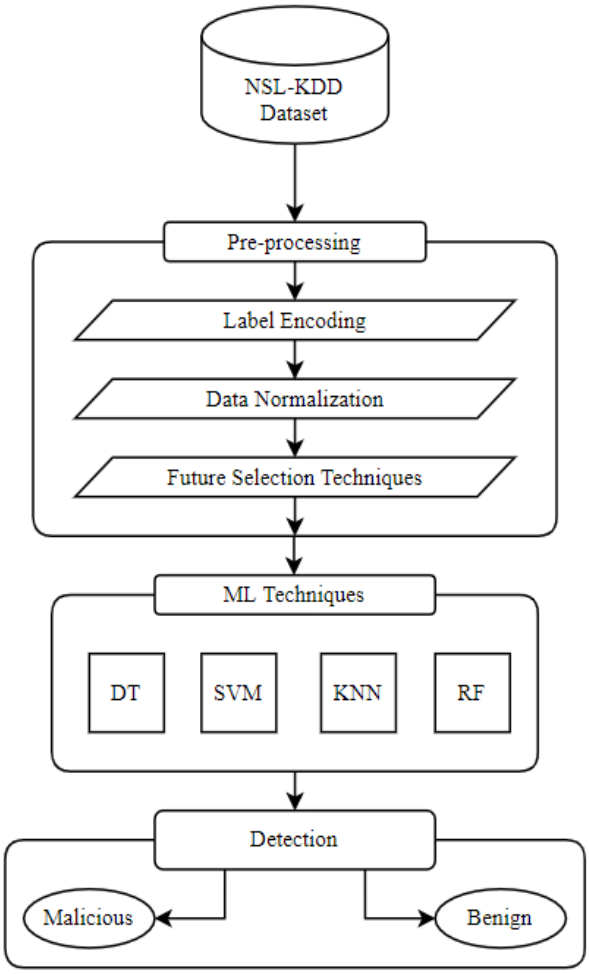


Figure 1
The proposed framework

3.3.1 Label Encoding

Label encoding converts categorical features to numbers in the NSL-KDD dataset. The NSL-KDD dataset contains three categorical features (protocol_type, service, and flag) that are converted to a number using the naming encoding method [30].

3.3.2 Data Normalization

The min-max normalization method was used to turn the numerical column values in the NSL-KDD dataset to a standard scale between 0 and 1 without distorting the value ranges. By using Equation (8) [31]:

$$Y = a - \min(a)/\max(a) - \min(a) \tag{8}$$

Where Y represents the normalized value, and *a* represents the original value.

3.3.3 Identifying Outliers in Regression (Cook's Distance)

Regression analysis helps to understand how variables in variables in groups of independent variables [32]. The aim of standardization is that the normal distribution of the values in our data set is symmetrical. Therefore, we will be able to detect whether the Mean, Mode, and Median data contain values close to each other [33]. We used Cook's distance, a successful method of detecting scaled changes in fit values, which is useful for identifying outliers (observations for predictive variables) of values in our dataset. An observation where Cook's distance is greater than three times the mean values may be an outlier. The Cook's distance of observation in Equation (9).

$$D_i = \frac{\sum_{j=1}^{n}(y_j - y_{j(i)})^2}{p^{MSE}} \tag{9}$$

Where:

- *yi* is the *j* th fitted response value
- *yj(i)* is the jth fitted response value, where the fit does not include observation *i*
- MSE is the mean squared error
- *p* is the number of coefficients in the regression model

3.3.4 Feature Selection Techniques

The proposed methodology was applied to the NSL-KDD dataset which has 41 features and one class attribute. After determining the importance value of each feature in the NSL-KDD dataset, the important features with a value greater than 0.05% were evaluated as shown in Table 2.

Table 2
The feature importance of the NSL-KDD dataset

No	Features Names	Features Importance	No	Features Names	Features Importance
1.	duration	1.14	**22.**	is_guest_login	<0.05
2.	protocol_type	13.44	**23.**	count	0.53
3.	service	1.93	**24.**	srv_count	0.66
4.	flag	56.96	**25.**	serror_rate	0.10
5.	src_bytes	0.74	**26.**	srv_serror rate	<0.05
6.	dst_bytes	0.68	**27.**	rerror_rate	<0.05
7.	land	<0.05	**28.**	srv_rerror rate	<0.05
8.	wrong fragment	1.38	**29.**	same_srv_rate	0.23
9.	urgent	<0.05	**30.**	diff_srv_rate	0.76
10.	hot	1.30	**31.**	srv_diff_host_rate	<0.05
11.	num_failed_logins	0.29	**32.**	dst_host_count	1.04
12.	logged_in	0.19	**33.**	dst_host_srv_count	0.80
13.	num compromised	<0.05	**34.**	dst_host_same_srv_rate	7.91
14.	root_shell	<0.05	**35.**	dst host_diff_srv_rate	0.40
15.	su attempted	<0.05	**36.**	dst host same_src_port_rate	1.73
16.	num_root	<0.05	**37.**	dst_host_srv_diff_host_rate	3.11
17.	num_file_creations	<0.05	**38.**	dst_host_serror_rate	0.06
18.	num_shells	<0.05	**39.**	dst_host_srv_serror_rate	0.18
19.	num_access_files	<0.05	**40.**	dst_host_rerror_rate	3.07
20.	num_outbound_cmds	<0.05	**41.**	dst_host_srv_rerror_rate	0.10
21.	is_host_login	<0.05			

The critical feature selection that affects the classification result is essential for a more accurate classification process. One of the most effective ways to select features is to use a DT technique to determine the feature's importance [34]. The feature importance property of the DT technique is beneficial for determining the importance of each feature and its effect on the classification result [35], [36]. This study has been developed to detect malicious or benign traffic in DDoS attacks. The NSL-KDD dataset was generated from the KDD dataset.

4 Results and Discussion

The four ML (RF, DT, KNN, and SVM) techniques are built using the Scikit Learn library, one of the powerful libraries used to build and implement ML techniques and data preprocessing in Python. With the Cook's distance method, the data found more than 3 times the threshold value were investigated and

necessary adjustments were made accordingly. Finally, the significance of the data was tested using both the T-test statistic and calculating the confidence intervals. It has been observed that our data are within the confidence interval.

The dataset used in this study was divided into 80% for training the models and 20% for testing. The performance of the four techniques was compared on only 25 features from the selected dataset using the feature selection technique. Among the techniques used, the RF technique achieved the highest accuracy, with 99.72%, superior to the rest of the techniques. While DT technique came in second with 99.51%, followed by the KNN technique with an accuracy of 99.42%, while the SVM technique achieved the lowest accuracy with 99.03%. Table 3 shows the performance comparison between the four ML techniques using 5 different performance measures. Also, Figures 3 to 6 show the confusion matrix of the four ML techniques.

Pre-processing and feature selection techniques are essential steps before implementing ML techniques. An exemplary implementation of data pre-processing and testing of critical features would increase the performance accuracy to approximately 45%. Therefore, in this study, the proposed model achieved promising results after selecting the important and influencing features and selecting the appropriate ML model.

Table 3

Performance evaluation of ML techniques

ML Techniques	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
DT	99.51	99.49	99.46	99.54	99.47
SVM	99.03	99.39	98.54	99.47	98.96
KNN	99.42	99.46	99.30	99.53	99.38
RF	99.72	99.84	99.56	99.85	99.70

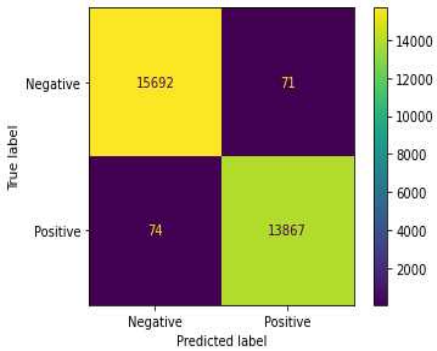


Figure 3

Confusion matrix of DT

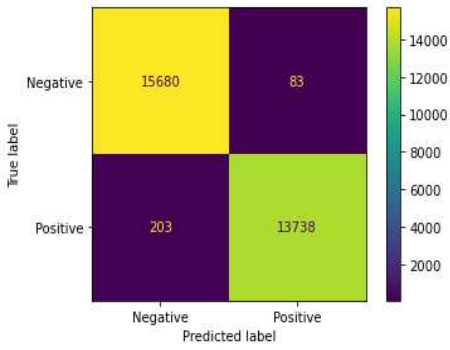


Figure 4

Confusion matrix of SVM

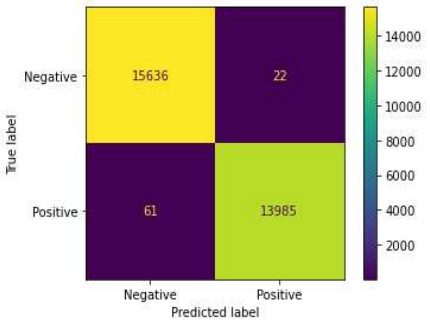


Figure 5
Confusion matrix of RF

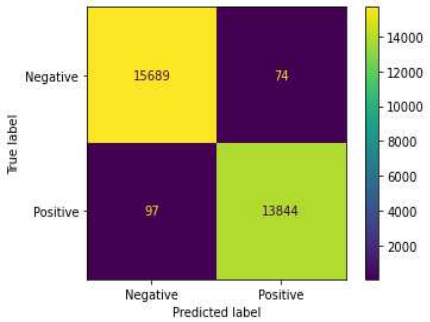


Figure 6
Confusion matrix of KNN

The RF technique used in this study was better than the results obtained from the RF technique used in other studies. In addition, the KNN, SVM, and DT techniques of this study performed better than KNN, SVM, and DT techniques. The recommended model, based on feature selection techniques and RF classifications, has achieved promising and reliable performance in the future to create identities based on ML techniques. In summary, it performed better than all studies mentioned in the study of the proposed model. Also, a higher accuracy ratio was obtained than all the studies mentioned in Table 4.

Table 4
Comparing the performance of the proposed model with related works

Ref.	Year	Dataset	ML Techniques	Best ML Technique	Best Accuracy
[15]	2019	CIC IDS	RF, LR, NB, KNN, Linear SVM, and LDA	RF	96.50%
[16]	2019	NSL-KDD	Ensemble model, MLP, SVM, KNN, and DT	Ensemble model	99.10%
[17]	2020	CICIDS2017 and NSL-KDD	AE+ SVM	AE+ SVM	96.36%
[18]	2020	CICIDS2017 and NSL-KDD	AE+DNN	AE+DNN	98.43%
[19]	2020	NSL-KDD	CNN and RNN	CNN	97.01%
[20]	2020	NSL-KDD	CNN	CNN	99.30%
[11]	2021	NSL-KDD	CNN, LSTM, and CLSTMNet	CLSTMNet	99.28%
[21]	2021	UGR16 and the UNSW-NB15, and KDD99	DNN	DNN	99.70%

[16]	2022	NSL-KDD	KNN, SVM, DT, and LR	LR	90.4%
Our Study		**NSL-KDD**	**RF, KNN, SVM, and DT**	**RF**	**99.72%**

Conclusions

There has been a significant increase in cyber-attacks, targeting organizations, institutions and even individuals. With the tremendous technological developments, the skill used by attackers, has increased and traditional IDS, can no longer detect sophisticated cyber-attacks. This required finding new, advanced tools to detect these destructive and expensive attacks. After the great successes of ML and DL techniques, in various fields, there have been many studies that use ML techniques in building IDS systems. This study presents an IDS based on feature selection techniques and ML techniques, for intrusion detection. The proposed model achieved promising results, as the RF technique achieved an accuracy of 99.72%, superior to other techniques in this work and related works. Having an intelligent system capable of detecting intrusion, helps significantly in maintaining the privacy and security of users. In this work, the focus is only on whether the traffic is malicious or benign. Future work could be developed to classify the different types of cybersecurity attacks. Classification accuracy can also be improved, by using ensemble methods, that combine more than one individual classifier.

References

[1] Dasgupta, D. et al.: Machine learning in cybersecurity: a comprehensive survey, *The Journal of Defense Modeling and Simulation*, 2022, Vol. 19, No. 1, pp. 57-106

[2] Al-Garadi, M. A. et al.: A survey of machine and deep learning methods for internet of things (IoT) security, *IEEE Communications Surveys & Tutorials*, 2020, Vol. 22, No. 3, pp. 1646-1685

[3] Salih, A. et al.: A Survey on the Role of Artificial Intelligence, Machine Learning and Deep Learning for Cybersecurity Attack Detection, *2021 7th International Engineering Conference "Research & Innovation amid Global Pandemic" (IEC)*, Erbil, Iraq, February 2021, pp. 61-66

[4] Zeebaree, S. R. et al.: Impact analysis of SYN flood DDoS attack on HAProxy and NLB cluster-based web servers, Indones, *J. Electr. Eng. Comput. Sci*, 2020, Vol. 19, No. 1, pp. 510-517

[5] Henry, A. Gautam and S.: Intelligent Intrusion Detection System Using Deep Learning Technique. In: *Computing, Communication and Learning: First International Conference, CoCoLe 2022, Warangal, India, October 27-29, 2022, Proceedings*, Cham: Springer Nature Switzerland, 2023, pp. 220-230

[6] Tong, W. et al.: A survey on intrusion detection system for advanced metering infrastructure, *In: Sixth international conference on instrumentation & measurement, computer, communication and control (IMCCC)*, IEEE, Harbin, China, July 2016, pp. 33-37

[7] Cloud Attack: Economic Denial of Sustainability (EDoS). Accessed: May 4, 2019 [Online] Available: http://www.elasticvapor.com/ 2009/01/cloud-attack-economic-denial-of.html

[8] AWS Said it Mitigated a 2.3 Tbps DDoS Attack, the Largest Ever. Accessed Jun. 30, 2020 [Online] Available: https://www.zdnet.com/ article/aws-said-it-mitigated-a-2-3-tbps-ddos-attack-the-largest-ever

[9] Academic Research Reports Nearly 30,000 DoS Attacks Per Day, Accessed Dec. 16, 2019 [Online] Available: https://www.corero.com/blog/853-academic-research-reports-nearly-30000-dos-attacks-per-day

[10] Halbouni, A. et al.: Machine Learning and Deep Learning Approaches for CyberSecuriy: A Review, *IEEE Access*, 2022, Vol. 10, pp. 19572-19585

[11] Issa, A. S. A. and Albayrak, Z.: CLSTMNet: A Deep Learning Model for Intrusion Detection, *In Journal of Physics: Conference Series*, 2021, Vol. 1973, No. 1, pp. 012244

[12] Abdullahi, M. et al.: Detecting cybersecurity attacks in the internet of things using artificial intelligence methods: A systematic literature review. *Electronics*, 2022, Vol. 11, No. 2, 198

[13] Salih, A. A. and Abdulrazaq, M. B.: Combining Best Features Selection Using Three Classifiers in Intrusion Detection System, *2019 International Conference on Advanced Science and Engineering (ICOASE)*, Zakho - Duhok, Iraq, 2019, pp. 94-99

[14] Kaggle, Accessed Dec. 16, 2019 [Online] Available: https://www.kaggle.com/datasets/hassan06/nslkdd

[15] Bindra, N. and Sood, M.: Detecting DDoS attacks using machine learning techniques and contemporary intrusion detection dataset. *Automatic Control and Computer Sciences*, 2019, Vol. 53, pp. 419-428

[16] Chavan, N., et al.: "DDoS Attack Detection and Botnet Prevention using Machine Learning," *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2022, pp. 1159-1163

[17] Das, S. et al.: DDoS Intrusion Detection Through Machine Learning Ensemble, *2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, Sofia, Bulgaria, 2019, pp. 471-477

[18] Kasim, Ö.: An efficient and robust deep learning based network anomaly detection against distributed denial of service attacks. *Computer Networks*, 2020, Vol. 180, pp. 107390

[19] Bhardwaj, A. et al.: Hyperband tuned deep neural network with well posed stacked sparse autoencoder for detection of DDoS attacks in cloud, *IEEE Access*, 2020, Vol. 8, pp. 181916-181929

[20] Al-Emadi, S. et al.: Using deep learning techniques for network intrusion detection, *In 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, Doha, Qatar, 2020, pp. 171-176

[21] Abu Al-Haija, Q., and Zein-Sabatto, S.: An efficient deep-learning-based detection and classification system for cyber-attacks in IoT communication networks, *Electronics*, 2020, Vol. 9, No. 12, pp. 2152

[22] Larriva-Novo, X. et al.: An IoT-focused intrusion detection system approach based on preprocessing characterization for cybersecurity datasets, *Sensors*, 2021, Vol. 21, No. 2, pp. 656

[23] Ghiasi, M. M. et al.: Decision tree-based diagnosis of coronary artery disease: CART model, *Computer methods and programs in biomedicine*, 2020, Vol. 192, pp. 105400

[24] Nhat-Duc, H. and Van-Duc, T.: Comparison of histogram-based gradient boosting classification machine, random Forest, and deep convolutional neural network for pavement raveling severity classification. *Automation in Construction*, 2023, Vol. 148, pp. 104767

[25] Kurani, A. et al.: A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting, *Annals of Data Science*, 2023, Vol. 10, No. 1, pp. 183-208

[26] Surinta, O. et al.: Recognition of handwritten characters using local gradient feature descriptors. *Engineering Applications of Artificial Intelligence*, 2015, Vol. 45, pp. 405-414

[27] Azrour, M. et al.: Machine learning algorithms for efficient water quality prediction, Model. Earth Syst. Environ., 2022, Vol. 8, pp. 2793-2801

[28] Tavallaee, M. et al: A detailed analysis of the KDD CUP 99 data set, 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, ON, Canada, 2009, pp. 1-6

[29] Fung, W. K. et al.: Influence diagnostics and outlier tests for semiparametric mixed models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 64, No. 3, pp. 565-579

[30] Jackson, E. and Rajeev, A.: Performance evaluation of different feature encoding schemes on cybersecurity logs, 2019 SoutheastCon, Huntsville, AL, USA, 2019, pp. 1-9

[31] Singh, D. and Singh, B.: Investigating the impact of data normalization on classification performance, *Applied Soft Computing*, 2020, Vol. 97, pp. 105524

[32] Kannan, K. S. and Manoj, K.: Outlier detection in multivariate data, *Applied mathematical sciences*, 2015, Vol. 47, No. 9, pp. 2317-2324

[33] Choetkiertikul, M. et al.: A deep learning model for estimating story points, *IEEE Transactions on Software Engineering*, 2018, Vol. 45, No. 7, pp. 637-656

[34] Issa, A. S. A. and Albayrak, Z.: DDoS Attack Intrusion Detection System Based on Hybridization of CNN and LSTM, *Acta Polytechnica Hungarica*, 2023, Vol. 20, No. 2, pp. 105-123

[35] Sugumaran, V. et al.: Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical systems and signal processing*, 2007, Vol. 21, No. 2, pp. 930-942

[36] Özalp, A. N. and Albayrak, Z.: Detecting Cyber Attacks with High-Frequency Features using Machine Learning Algorithms, *Acta Polytechnica Hungarica*, 2022, Vol. 19, No. 7, pp. 213-233

Pilot Signal Removal, Digital Signal Processing Algorithm and it's Practical Implementation

Tibor Wührl

Óbuda University, Kandó Kálmán Faculty of Electrical Engineering
Bécsi út 96/b, H-1034 Budapest, Hungary
wuhrl.tibor@kvk.uni-obuda.hu; ORCID: https://orcid.org/0000-0002-7522-3511

Abstract: In this paper, an algorithmic development, using an efficient wave, digital signal processing method, is presented. The developed algorithm can also be used in practice. The algorithm is obtained by transforming a passive reference network, the passive reference network is described by voltage waves. After the theoretical summary, the steps of design and design review are presented through a concrete practical example. The algorithm is guaranteed to be stable and limit-cycle free.

Keywords: UIC pilot signal; DSP algorithm; Digital Signal Processing; Wave Digital Filtering

1 Introduction

The International Union of Railways (UIC) defines the specifications for a number of railway communication systems and constituents.

The UIC codex 751-3, 4th edition 2005 [1] gives the "Technical specifications for analogue train radio systems in international service". The specifications of analog systems cannot be called new, but due to compatibility requirements, old system specifications must also be taken into account when designing new system components. The new system components are preferably made with digital solutions at the technical level of our time and thus also in the case of analog systems. Circuits implemented in digital circuits operating in analog environments take a sample from the analog signal then quantize and encode them, i.e., digitize them. Output samples are calculated from the digital samples according to the signal processing algorithm in real time [2]. The DSP algorithm defines the calculations and their order. The creation of DSP algorithms is a mathematical and electrical engineering task, during which several mathematical transformations are required.

In every step of the design of the signal flow diagram, we have to take care of the stability, i.e., that the developed algorithm does not excite and is free of limit cycles [3]. This can be guaranteed by maintaining passivity throughout [4]. In the case of wave-digital filter design, our reference circuit is composed of passive components (resistor, capacitor, inductance, and lossless interconnect components) [3], but the stability conditions can only be met if the implementation is not included any faults. We use a linear model throughout the design of the signal flow diagram, however, all of the components in the model are lossless. In the implementation of the algorithm (FPGA, or signal processing processor design), the operating environment is not linear due to the finite number of bits. I treat the granular nonlinearity resulting from quantization as well as the overflow separately. MATLAB® simulation is used to handle problems and detect individual faults [5].

1.1 Transformation System, Design Steps of DSP Algorithm

The DSP algorithm is built on different planes. Transformations give connections between each plane. The mathematical relationships between the planes are important elements of the design steps. All of this is summarized in the following figure:

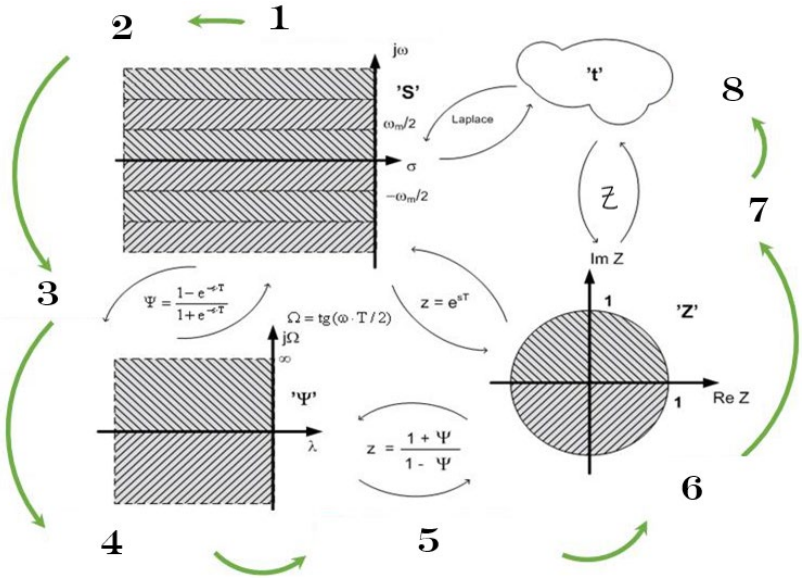


Figure 1
System of Domains and Transformations

The step marked '1' in the figure above is to specify the requirement on the complex frequency plane. In the quasi-stationary case, the real part of the complex frequency is zero, so the requirement is defined with respect to jω, as an independent variable.

Figure 2 shows the requirement for a low pass filter. The requirement for high-pass, band-stop, and band-pass filters can be converted to a low-pass requirement after so-called type transformation [6], that will not detailed in this article.

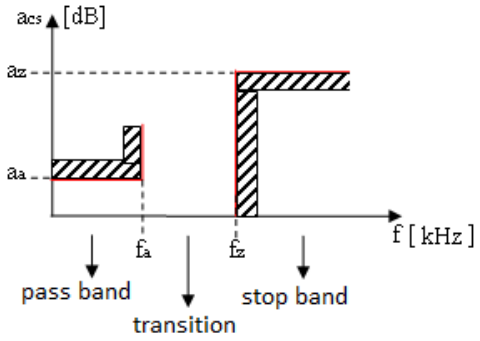


Figure 2
Low-pass filter criteria in frequency domain

In the next step, we select a sampling frequency for the digital signal processing system that matches the spectrum of the signal to be processed. The value of the sampling frequency is given in the Sannon's theorem:

$$fm \geq 2 \cdot f\,\max \tag{1}$$

As the result of sampling, the 's' complex frequency plane and its jω axis become periodic. The alias components are displayed. This means that the requirement in Figure 2 also becomes periodic in the + ∞ and -∞ range. The band periodicity makes it impossible to use well-proven mathematical methods in the design of circuits and algorithms. Step 3, of Figure 1, eliminates periodicity with the following relationship:

$$\Psi = \frac{1 - e^{-sT}}{1 + e^{-sT}} \tag{2}$$

Equation (2) makes the plane Ψ free of physical dimensions. Several DSP literatures [7] use a different transformation and define the plane Ψ denoted as a distorted but physically dimensional plane. Exemption from the physical dimension has a number of advantages that are an advantage in the design steps.

The Ψ plane is of complex frequency nature, the real part of that is denoted by λ, while the complex component is denoted by Ω, so the independent complex variable is as follows:

$$\Psi = \lambda + j\Omega \tag{3}$$

Important features of the transformation (2) are that it transforms the 's' plane origin into the 'Ψ' plane origin, as well as an axis-bearing transformation, i.e., moves the jω axis to the jΩ. For σ = 0, transformation (2) will have the following form (4):

$$\Omega = tg(\omega \cdot \frac{T}{2}) \tag{4}$$

Of course, as we have established the physical lack of dimensions in the plane Ψ, this is now also true for the Ω axis. Transformation (4) eliminates the periodicity that became periodic after sampling (Figure 3) (3rd step of Figure 1):

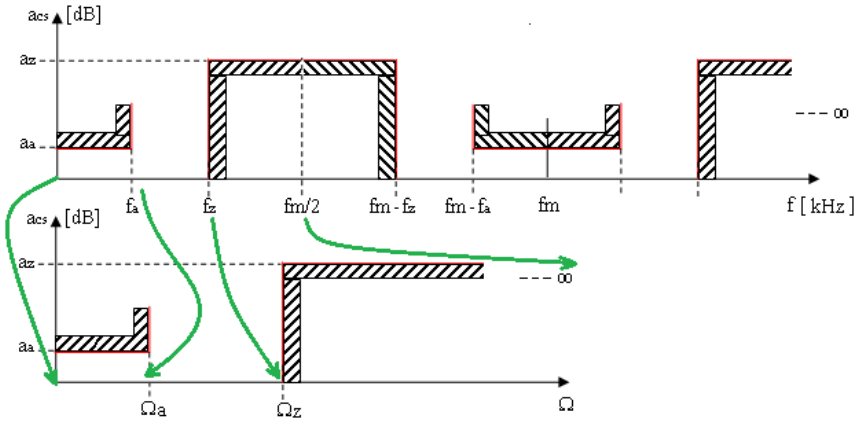


Figure 3
Elimination the periodicity of requirement

Figure 3 shows that half of the sampling frequency is transformed from the 's' plane, to the + ∞ point in plane Ψ.

The fourth step in Figure 1 is designing an analog reference filter in the Ψ domain. In this domain we can use all well-known analog passive filter design methods. These are not intended to be presented in this article.

The fifth step in Figure 1 is more complex.

Before beginning the transformation, it is advisable to redraw the analog reference circuit. The following figure shows an example of how to connect the individual components per gate.

Based on Figure 4, the individual building elements were highlighted as passive components, as well as the passive network components that connect them. Passive connections can be parallel or serial.

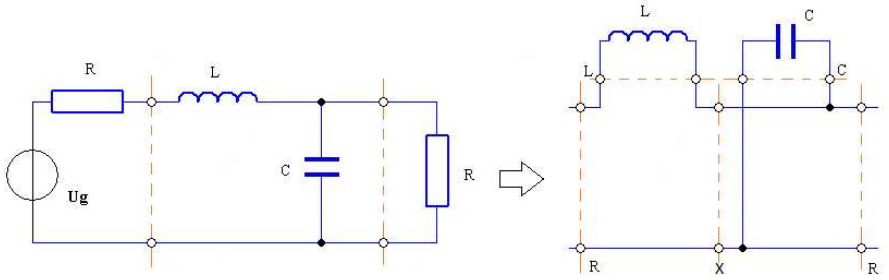


Figure 4
Redrawing the reference filter in "Ψ" domain (an example)

The transformation of the components related to the gate coupling [4], is discussed in the next subsection due to its significance.

1.2 Transformation of Components with Voltage Waves

As described in the previous section, the reference circuit is composed of passive components only. By separating these passive components from each other and coupling them as gates, we can obtain separately transformable elements. Our goal is to transfer these components to the "z" plane. During the transformation, care will be taken to maintain passivity. Each building block is described by voltage waves (5).

$$\begin{aligned} a(t) &= u(t) + i(t) \cdot R \\ b(t) &= u(t) - i(t) \cdot R \end{aligned} \tag{5}$$

In the context of (5) above, a (t) is the voltage wave passing through the gate, while b (t) is the reflected wave. "R" is the gate resistance parameter specific to the gate. The "Ψ" plane does not contain physical dimensions, so the gate resistance parameters are just mere numbers in Figure 4 marked "R", "L", "C" and "X". The traveling and reflected waves can, of course, be interpreted in all the planes defined in Figure 1, for example in the "Ψ" plane (6):

$$\begin{aligned} A(\Psi) &= U(\Psi) + I(\Psi) \cdot R \\ B(\Psi) &= U(\Psi) - I(\Psi) \cdot R \end{aligned} \tag{6}$$

Thus, we interpreted propagating and reflected waves between the gates, and assigned a gate resistance to the gates.

In describing the transformations, I have shown that the plane Ψ is a plane free of physical dimensions. We will also take advantage of this, since the gate resistance R has only a numerical value. The same is true for capacitor "C" and inductance "L".

Let us examine how the capacitor "C" defined in the plane Ψ can be transformed into the plane "z":

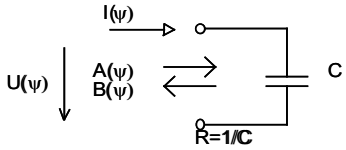


Figure 5
"C" components on "Ψ" domain

The relationship between current and voltage on the capacitor can be determined by the following equation:

$$U(\psi) = I(\psi) \cdot \frac{1}{\psi \cdot C} \tag{7}$$

The voltage and current can be described by the voltage propagation and reflected waves in the plane Ψ, which are derived from (6):

$$U(\psi) = \frac{A(\psi) + B(\psi)}{2}$$

$$I(\psi) = \frac{A(\psi) - B(\psi)}{2 \cdot R} \tag{8}$$

Substituting the current and voltage values expressed by voltage waves (8) into the Equation (7), we get:

$$\frac{A(\psi) + B(\psi)}{2} = \frac{A(\psi) - B(\psi)}{2 \cdot R} \cdot \frac{1}{\psi \cdot C} \tag{9}$$

After choosing the gate resistance condition R = 1 / C, our relation (9) will have the following form:

$$B(\psi) = A(\psi) \cdot \frac{1 - \psi}{1 + \psi} \tag{10}$$

Applying the relation marked in step 5 at (Figure 1) of our transformation system we get the following:

$$B(z) = A(z) \cdot z^{-1} \tag{11}$$

Equation (11) gives the building element transformed to the "z" plane (Figure 6). Here we must not forget that this result was obtained in connection with the choice of the gate resistor R = 1 / C.

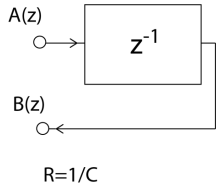


Figure 6
Transforming the "C" Components to "z" domain with voltage waves

Let us examine how the inductance "L" defined in the plane Ψ can be transformed into the plane "z". With a similar line of reasoning and choosing the gate resistor R = L, the result will be:

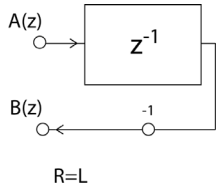


Figure 7
Transforming the "L" Components to "z" domain with voltage waves under condition R=L

The parallel connecting network part can be described using the Kirchoff's laws. In the simplest case, two components are connected in parallel, but of course the connection can also be done by connecting a "n" gate. Connecting network components are called adaptors [4] [8].

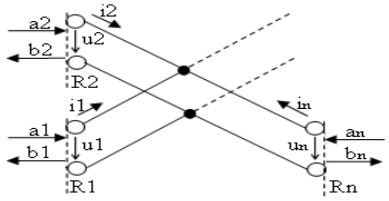


Figure 8
Parallel connection model

Excluding derivations, the general wave equation for the parallel connection is as follows [3]:

$$b_v = (\gamma_1 \cdot a1 + \gamma_2 \cdot a2 + ... + \gamma_n \cdot a_n) - a_v \tag{12}$$

In the context of (12), G is the gate conduction parameter which is the reciprocal of the gate resistance R. In the context, "n" is the total number of gates, and the index "v" is the number of the gate to be characterized.

With the gate guidance parameter, the reflection factor of a given gate is given by the following relation [3]:

$$\gamma_\nu = \frac{2 \cdot G\nu}{\sum_{k=1}^{n} G_k} \tag{13}$$

The initial figure for the serial connection is illustrated in the following figure:

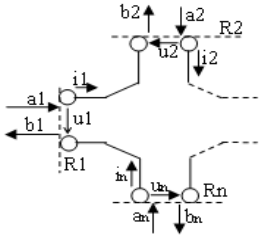


Figure 9
Serial connection model

Excluding derivations, the general reflected wave of the series connection at the ν^{th} gate is as follows [3]:

$$b_\nu = a_\nu - \gamma_\nu \cdot (a1 + a2 + ... + a_n) \tag{14}$$

The reflection factor of equation (14) with the gate resistance parameters R will be as follows [3]:

$$\gamma_\nu = \frac{2 \cdot R\nu}{\sum_{k=1}^{n} R_k} \tag{15}$$

It can be seen from (12) (13) and (14) (15) that the adaptor components are memory - free, the interconnection of the components is loss - free, and can be considered passive under linear operating conditions.

1.3 Simulation Method before Realization of Algorithm

A linear model of the DSP algorithm was generated in the “z” plane. This is well suited for thinking in the frequency range.

The operation of the DSP algorithm, i.e., the input pattern at the sampling intervals (T) and the state variables storing the previous effects, determine the output pattern of the given phase. The signal flow diagram to be implemented will thus be a task scheduled at “T” intervals. This matches the DSP flowchart drawn

in the time range. For the above reasons, the simulation and the effect of nonlinearities are investigated in the time domain. Observation and model creation in the time domain are used in many fields also [15].To do this, the DSP algorithm generated in the "z" plane should be transformed into the time domain (see Fig. 1, step 7).

The tests are performed in the time domain with MATLAB® simulation, the advantage of which is that it can give results close to the actual implementation. The purpose of the primary results is to support the fulfillment of the filter requirement. Then, within the limits of the number representation accuracy provided by MATLAB® [9], The simulation environment and the obtained result are considered as a study under linear conditions. The purpose of this simulation is to shed light on possible design errors. The Kronecker delta signal as the input, excitation signal was used. This signal assumes a value of "1" at the time of first sampling, so that the spectrum of the signal is "1" for all possible spectral components, although for a period of "dt". Of course, the examined filter modifies the amplitude and phase position of the spectral components of the input signal according to its transmission function. The applied simulations are even used in the time domain analysis of mechanical systems too [17].

In the time domain, the response to the Kronecker delta signal will be the weight function, by whose spectral analysis we can determine the transfer function. For the simulation result, was expect that the transfer function fulfills the expectation for all parameters of the filter requirement given for the initial condition (see in Figure 2). If this was not the case, then we made a mistake in one of the points 1 to 7 of the design (see in Figure 1), so it is necessary to check and correct the individual steps.

After the successful simulation in the environment considered as linear conditions, was consider it necessary to investigate the nonlinear effects caused by the finite number representation. This also requires running the MATLAB® time domain simulation [5]. Two types of tests should be performed with a test signal that matches the typical conditions of use of the algorithm.

The first aspect of the study is the overflow analysis. Here, it must be determined whether there are points in the implementation (FPGA or microcontroller) that the algorithm contains overflow-prone nodes for the available number representation (case of given bit number and representation mode). Based on the results of the discovery of such nodes, the control of these points can be reduced by so-called scaling [10]. During scaling, the risk of overflow can be eliminated or reduced by replacing some network component with a linear equivalent [10]. This intervention ultimately improves the dynamic range of the algorithm.

The second nonlinearity test assumes a more accurate knowledge of the actual algorithm implementation. Now we need to know exactly the quantization points that are forced to be used in the algorithm during implementation. These typically occur at the output of multipliers, since in the case of a fixed-point representation,

the result swells to 2 times "n" bits by multiplying the "n-bits" represented numbers with each other. Bit count truncation (rounding, truncation) occurs as a noise generating generator. In addition to signal-to-noise degradation, excess signal energy appears in the algorithm, compromising the passivity criterion.

2 Pilot Signal Remover Filter

After the theoretical introduction, the design and feasibility are presented through a specific task. In formulating the requirement, the goal is to remove a pilot carrier signal mixed to an audio frequency signal [1]. A band-stop filter with a sufficient slope is suitable for this. The stop band of the band should be as narrow as possible in terms of the speech signal, but the unpleasant effect of the components resulting from the frequency inaccuracy of the pilot signal and possible switch-off phenomena is contrary to this requirement. It is important to maintain a “balance” in the requirement. The LWDF structure was chosen to solve this problem. [10].

2.1 Structure of Lattice Wave Digital Filter []

The LWDF filter consists of two branches, a lower branch and an upper branch. A Figure 10, in the illustrated structure, S’ and S” are called the “all-pass components”.

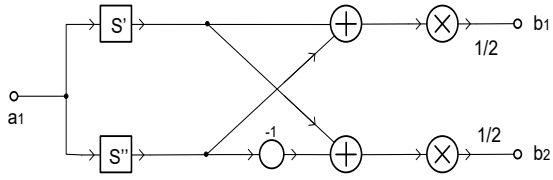


Figure 10
Applied LWDF structure

The absolute value of the transmission function of the all-pass component is a constant one regardless of the complex frequency. This means that these terms do not change their gain as a function of frequency, but do phase shift only. At the frequencies at which the phase shifts of S' and S” are the same and the two signals are summed, a passband is obtained. At frequencies where the phase difference between the phase rotations of the two all-pass parts outputs is 180 °, a closing band is obtained due to signal quenching. Output “b2” gives the dual pair of the result of “b1”.

For example, a first-order all-pass function can be formed with a circulator and a capacitive element:

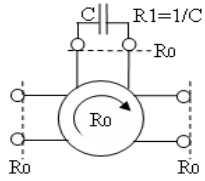


Figure 11
First order all-pass component at "Ψ" domain with circulator and capacitor

In practice, an all-pass component can be made by connecting a two-gate adaptor and a reactant component. These components were described in the previous chapter.

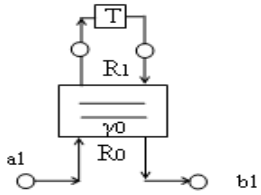


Figure 12
First order all-pass component at "time" domain with two port adaptors

The input of the first-order all-pass section is now the input of the forward wave (a1) of gate number 1, while the output is the reflected wave point (b1) of the same gate. The γ0 multiplication factor of adaptor is now not calculated from the indicated gate resistances, but is defined by the following relation (12):

$$\gamma_0 = \frac{1 - B_0}{1 + B_0} \tag{12}$$

Figure 13 illustrates the n^{th} degree LWDF structure built from all-pass components.

2.2 Pilot Remover LWDF

According to UIC Code 751-3, 4^{th} edition of 2005 [1], the audio frequency speech signal broadcast on the radio sits on a 2800 Hz harmonic pilot carrier. The pilot carrier is in the audible range during reception, so it is absolutely necessary to remove it, in order to understand the transmitted speech.

Samples are available from the analog signal at intervals of 125 μs, i.e., the sampling frequency is 8000 Hz.

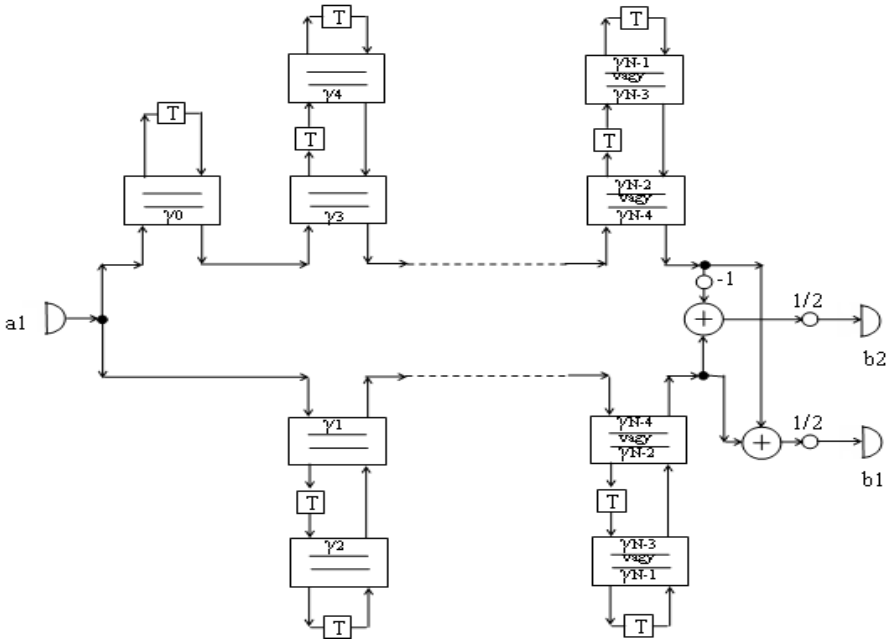


Figure 13
n^{th} order LWDF structure

At the frequency of 2800 Hz, the minimum requirement is 30 dB attenuation, therefore a starting demand of 40 dB is assumed for the approximation (This means an attenuation margin of 10 dB.). The approximation tasks are solved with the MATLAB® program. Using the MATLAB toolbox [11], the specified parameters are as follows without explaining the execution steps:

- fs=8000 Hz
- Center frequency = 0.35 (relative from fs)
- Bandwidth = 0.09 (relative from fs)
- Filter type: band stop

The approximation method is Butterworth.

The reflection factors of the individual adapters are given the following numerical values:

$\gamma 1 = -0.88701$

$\gamma 2 = -0.58779$ (13)

The above filter coefficients can be directly applied to the all-pass components shown in Figure 12 and Figure 14.

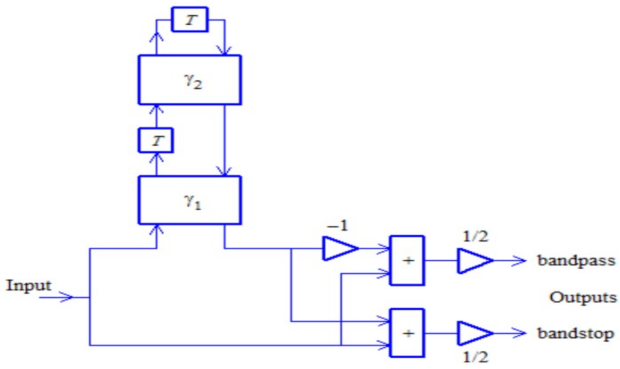


Figure 14
UIC pilot signal remover LWDF structure, designed with MATLAB® [11]

Under linear conditions, this would give a satisfactory result, but due to the quantization and overflow problems caused by the finite number of bits, the scaling of the structure [10] becomes necessary.

2.3 Scaling of the Pilot Remover LWDF

When scaling WDFs, we use the linear equivalents of each building block. In linear conditions, this has no effect. In a finite-bit implementation environment, the linear equivalents [8], on the other hand, behave differently.

High-frequency theoretical foundations [14] are often used in LWDF design.

After scaling, the filter coefficients were changed (14):

$$g1 = 1 + \gamma 1 = 0.11299$$

$$g2 = 1 + \gamma 2 = 0.41221 \quad (14)$$

After scaling the filter, we should check that it is working properly.

The detailed signal flow diagram of the scaled DSP algorithm is illustrated in Figure 15.

For verification, a time-domain simulation was performed on the MATLAB® platform.

A lot of literature deals with the description and simulation of time series [12], [13]. The planned algorithm is also now expediently simulated in the time domain.

In Figure 15, the nodes calculated in the time-domain simulation are marked with capital letters A - G.

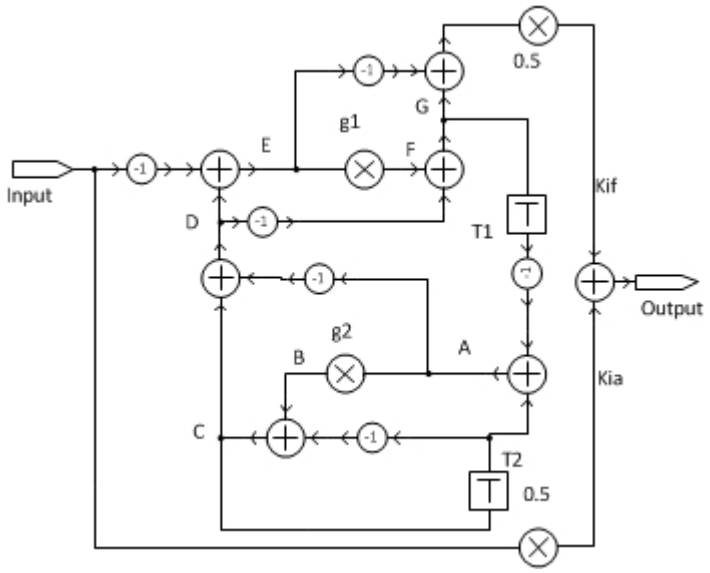


Figure 15
Detailed algorithm of UIC pilot signal remover - optimized

The order of the time-domain simulation fits well with the implementation with the signal processing processor; therefore the "m" source code has also been published. It can be downloaded from this link: https://bit.ly/UIC_filter1

After running the code, one of the results is shown in Figure 16.

Based on Figure 16, it can be seen that the response (weight function) to the Kronecker delta excitation signal has a stable decay. This result confirms the stability of the filter. Algorithm stability is also highly significant in the field of sound and 2D image processing [17]. The Fourier transform of the Kronecker delta is a constant “one”. This means that during one sampling time unit (ideally “dt”), all spectral components are present in the excitation signal with unit amplitude. For this reason, it is also advisable to examine the spectral image of the response function, which in this case is the transfer function of the filter circuit. See Figure 17.

It can be seen from the transfer function that the filter has significant attenuation at a frequency of 2800 Hz. According to the result, the 2800 Hz harmonic pilot signal will be attenuated by approximately 80 dB. This value is much better than expected in the requirement.

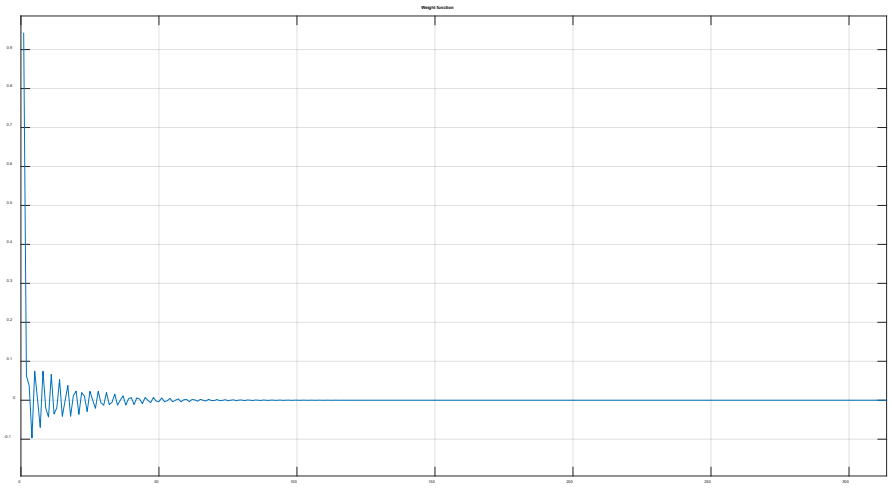


Figure 16
UIC pilot signal remover filter simulation result in time-domain

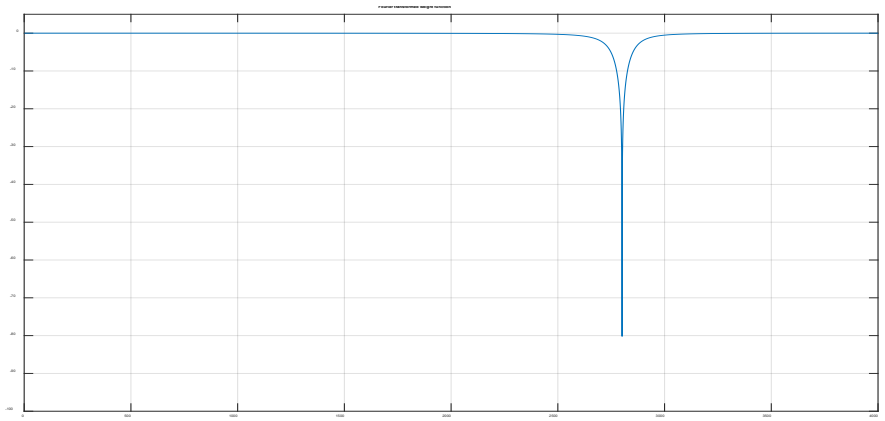


Figure 17
UIC pilot signal remover filter simulation result in frequency domain (Transfer function)

2.4 Questions of Realization

We kept the passivity in some steps of the design, thereby ensuring zero input signal stability of the DSP algorithm. The algorithm was also scaled. In this step, the absolute value of the constant multiplier value of the multiplier elements is kept below 0.5. The use of linear equivalents in the algorithm does not endanger its stability, but gives better dynamic range.

For better dynamics, the output 0.5 multiplier is also moved in front of the adding element (compare Figure 14 with Figure 15):

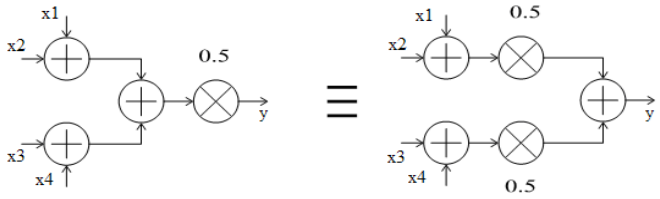


Figure 18
Output node conversion of UIC pilot signal remover

The linear equivalence in Figure 18 can also be seen mathematically.

When implementing DSP algorithms, numbers are represented using a finite number of bits. The effect of the finite number of bits becomes apparent when the algorithm is implemented cost-effectively, that is, by incorporating the actually necessary hardware components. For real-time DSP solutions, we often use fixed-point number representation, for example according to the IEEE1.15 or IEEE1.31 format. In the case of such solutions, the loss of even a few bits from the useful operation results in a low-quality solution. The finite number of bits means a perceptible quantization error and this error affects the operation of the algorithm in a non-linear manner. In addition to the appearing quantization noise, a bigger problem is that the value of the sample at certain nodes of the algorithm is either increased or decreased by the quantization effect. With this DSP algorithm, the pseudo energy of the signal represented by the patterns generated in each node can also increase. An increase in energy threatens passivity. In case of loss of passivity, we lose the guarantee of zero input stability.

An example of this is presented, in which mathematical rounding is applied to the output of the multiplier components of the scaled UIC filter based on the IEEE1.15 representation in the simulation. After multiplication, the number represented by IEEE1.31 is converted to IEEE1.15 format by mathematical rounding. Rounding in the simulation is done with the following MATLAB® code:

```
function [q]=q1(x,n)
            q=(round(x*2^n))/2^n          (15)
```

The "n" represents the number of bits between the radix point and the LSB, "x" is the number to be rounded, which must be entered in decimal form. The function returns the value “q”. We can see the impulse response in Figure 19.

The weight function of the algorithm containing mathematical rounding goes into an infinite limit cycle from the 100th beat, marked with red line in the Fig 19. With the mathematical rounding, we violated the passivation conservation rule, so we lost the zero-input signal stability guarantee.

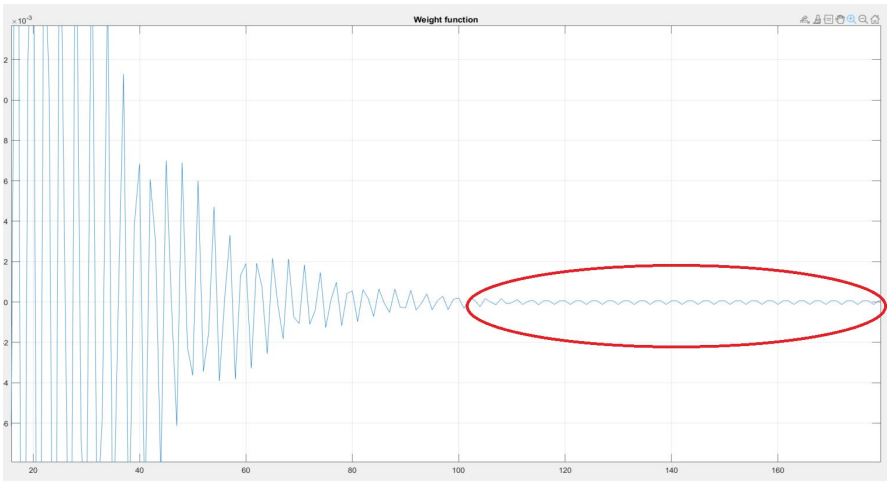


Figure 19
Infinite limit cycle in weight function

Of course, this phenomenon is also present in the case of floating-point number representation using a high number of bits, but the hardware, which is significantly oversized for the task, masks the errors. The existence of the error carries a source of danger, since nothing guarantees when a small-amplitude limit cycle will degenerate into an uncontrolled, high-amplitude excitation.

The passivity can also be maintained in the implementation of the DSP algorithm, if the quantization is performed with absolute value truncation. In this case, the samples representing the signal are pushed towards the zero point at each quantization point, that is, the signal energy is reduced in each case.

Conclusions

This paper illustrates the process efficiently designing an effective DSP algorithm and how to guarantee its' stable operation. The UIC pilot filter presented herein, can be successfully implemented, with minimal hardware requirements.

Acknowledgement

This work was supported by Óbuda University.

I thank the management of Óbuda University for supporting my work and I thank the Acta Polytechnica Hungarica journal for publishing my article.

References

[1] International Union of Railways: UIC codex 751-3 2005./4

[2] T. Wührl: DSP algoritms OE – KVK 2116, Budapest, 2014

[3] T. Wührl: Wave Digital Filtering, OE-KVK 2073. Budapest, 2010

[4] A. Fettweis: Wave Digital Filters: Theory and Practice, Proceedings of the IEEE, Vol. 74, No. 2, pp. 270-327, 1986

[5] T. Wührl:P Introduction in to the MATLAB: Telecommunication and DSP, OE-KVK 2071, Budapest, 2010

[6] K. Géher, J. Solymosi: Linear Circuit Design, Budapest, 1992

[7] R. Lyons: Understanding Digital Signal Processing, Prentice Hall, Upple Saddle River, NJ, 2001

[8] A. Fettweis, K. Meerkötter: On Adaptors for Wave Digital Filters, IEEE Transactions on Audio, Speech, and Signal Processing, Vol. ASSP-23, No. 6, pp. 516-525, 1975

[9] MathWorks: Representation of Numbers in Matlab, https://uk.mathworks.com/matlabcentral/fileexchange/33992-representation-of-numbers-in-matlab

[10] L. Gazsi: Explicit Formulas for Lattice Wave Digital Filters, IEEE Transactions on Circuits and Systems, Vol. 32, No. 1, pp. 68-88, 1985

[11] M. Tockner, H. G. Brachtendorf, M. Steiger: Design of Wave Digital Filters with the TU Delft Toolbox, Proceedings of 2021 16th International Conference on Telecommunications (ConTEL), pp. 18-22, 2021

[12] 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

[13] C. Andreeski, D. Mechkaroska: Modelling, Forecasting and Testing Decisions for Seasonal Time Series in Tourism, Acta Polytechnica Hungarica, Vol. 17, No. 10, pp. 149-171, 2020

[14] D. Singh, A. Shukla: Manifold Optimization with MMSE Hybrid Precoder for Mm-Wave Massive MIMO Communication, Romanian Journal of Information Science and Technology, Vol. 25, No. 1, pp. 36-46, 2022

[15] C. Pozna, R.-E. Precup: Aspects Concerning the Observation Process Modelling in the Framework of Cognition Processes, Acta Polytechnica Hungarica, Vol. 9, No. 1, pp. 203-223, 2012

[16] N.-A. Serban, R. Hobincu: Computer Vision Algorithm for Detecting Resistor Color Codes, Romanian Journal of Information Science and Technology, Vol. 24, No. 3, pp. 321-333, 2021

[17] A.-I. Szedlak-Stinean, R.-E. Precup, E. M. M. Petriu, R.-C. Roman, E.-L. Hedrea, C.-A. Bojan-Dragos: Extended Kalman filter and Takagi-Sugeno Fuzzy Observer for a Strip Winding System, Expert Systems with Applications, Vol. 208, paper 118215, 2022

A Simulation System for Testing Side Crashes, in Non-Traditional Seating Positions, for Self-Driving Cars

Laszlo Porkolab, Istvan Lakatos

Audi Hungaria Faculty of Automotive Engineering, Department of Road and Rail Vehicles, Széchenyi István University, Egyetem tér 1, H-9026 Győr, Hungary
porkolab.laszlo@sze.hu, lakatos@sze.hu

Abstract: Historically, vehicle safety reflects the current state of the art and new innovations will continue to make our cars even safer in the future. Since its invention, the car has enjoyed a unique, triumphal procession. Safety plays a central role in the development of a car model today. The number of accidents has also risen, in line with the growth in traffic. In addition to carelessness or distraction at the wheel, the most common causes of accidents are excessive speed, risky maneuvers and disregard of traffic rules. The introduction of the speed limit on rural roads and the obligation to wear seat belts, were particularly important positive milestones. Two different methods are also used to check the effect of the individual technical options and safety, in the event of accidents. First are the crash tests. Here, an accident situation is simulated in practice, under realistic conditions. The other solution is the simulation. The Finite Element Method, behind this term, lies the virtual calculation of various consequences of an accident, on the basis of mathematical differential equations. The degree of deformation of various components or the entire car, as a whole, is examined by calculation.

Keywords: crash test; occupant safety; finite element method; electric vehicles; autonomous driving

1 Introduction

Safety technology played almost no role in the first automobiles that rolled down our streets at the beginning of the 20th Century. However, the increase in the power of engines with higher speeds and the increase in traffic have gradually led to the fact that cars are becoming more and more extensively equipped in this area. From the beginning to the present day, vehicle safety also reflects the current state of the art. New innovations should continue to make our cars even safer in the future. Since its invention, the car has enjoyed a unique triumphal procession. Not only the vehicles themselves, but also the necessary infrastructure, has been consistently developed. Safety plays a central role in the development of a car model today.

Today, the number of accidents is declining in many places, despite the steadily increasing volume of traffic. In addition to restrictions due to legal regulations, the modern safety equipment of the vehicles is also responsible for this. While the vehicles did not have a proper body at the beginning, this aspect became more important from around 1920. In addition to greater comfort and a wide range of options for shaping the appearance of the car, the body was also used specifically to improve stability. Above all, the protection of the occupants was initially in the foreground. Different systems emerged. From floor plates or frames on which the body is built, to lattice frames and the "space frame", from AUDI to the self-supporting body (monocoque), these are still used today, depending on the vehicle type. In the post-war period, the bodies were increasingly produced by the car manufacturers themselves and the principle of the self-supporting body gradually prevailed in passenger cars. The chassis and body form a single unit and the overall structure offers great stability to various loads (shearing or torsional loads). A body with a crumple zone was used for the first time in 1959 on the Mercedes Benz W111. It ensured that the impact energy in the event of an accident was initially absorbed by the body, so that the forces on the occupants were less. Today, the crumple zone in a vehicle has different tasks. First of all, the bumpers, for example, ensure that the energy can be absorbed in very small collisions and other vehicle parts are spared damage. In addition, the front part of the vehicle in particular is designed in such a way that the impact energy is reduced at medium speeds and the rear of the car remains as intact as possible. The latter in particular is designed to be particularly stable in order to minimize the risk of injury.

The number of accidents has also risen for a long time in line with the growth in traffic. In addition to carelessness or distraction at the wheel, the most common causes of accidents are excessive speed, risky maneuvers and disregard of traffic rules. Technical defects are also a reason. Due to the consistent upgrading of vehicles with the appropriate safety technology, the number of accidents has been declining since around 1970, despite the further increase in traffic. In addition, new rules and regulations gradually began to have an effect. The introduction of the speed limit on rural roads and the obligation to wear seat belts were particularly important milestones.

Two different methods are also used to check the effect of the individual technical options and safety in the event of accidents. First are the crash tests. Here, an accident situation is simulated in practice under realistic conditions. The vehicles are equipped with extensive sensor technology and the crash is documented on video from a wide variety of perspectives. "Crash Dummies" are used as the vehicle occupants. They are equipped with sensors and can provide information about the degree of an injury. In 1996, the testing company Euro NCAP (European New Car Assessment Program) was founded by the British Department for Transport and other European organizations quickly followed suit. The increasing publication of results from crash tests to inform consumers has led

to vehicle safety becoming a greater focus in the development of new vehicles. Every year there are innovations in the assessment, so that the results are only really comparable within a year. The other solution is simulation. Finite Element Method (FEM), behind this term, lies the virtual calculation of various consequences of an accident, on the basis of differential equations. The degree of deformation of various components or the entire car as a whole, is examined by these calculations.

The possibilities of improving safety through driver assistance systems are becoming more and more extensive. By monitoring the traffic area around the vehicle with the help of a wide variety of sensors, they can be used to prevent accidents and driving errors caused by fatigue or carelessness. The more extensive the technology becomes, the more the vehicle approaches autonomous driving. As more and more modern technology is installed in vehicles, they have become heavier in recent years. To counteract this, innovative materials are increasingly being used for construction. Aluminum, magnesium, but also fiber composite materials help to reduce the weight. Depending on the intended use and type of component, the best possible material can be selected and, in the case of hybrid construction, assembled into a body. The modular design that is predominantly common today ensures that each individual part offers maximum safety thanks to its specific design and contributes to the overall stability of the body.

The progressive developments in the field of artificial intelligence and the first practical projects are currently testing the use of self-driving vehicles in road traffic. The current state of technology already makes this possible, but completely different questions arise here. In various emergency situations, the vehicle system may have to make decisions that affect people's lives. The example here is a situation in which the vehicle is moving towards two people and only one of them could be saved, by an evasive maneuver. In order to take the various security aspects into account, there are also ethical criteria that are difficult to implement with the automatic control of artificial intelligence. The question of liability in the event of a simple accident has also not yet been clarified. Self-learning systems and increasingly sophisticated mathematical calculation models should ensure maximum security here. Before autonomous driving can be integrated into everyday life, however, a number of open questions must first be clarified. Adequate protection against unauthorized access to the systems from outside is also not yet fully developed.

2 Vehicle Occupant Safety

Vehicle occupant safety refers to the combination of active and passive vehicle safety. While active vehicle safety helps to avoid accidents, passive vehicle safety deals with measures to minimize the consequences of an accident.

Active safety systems record driving conditions using sensors and intervene to support driving operations. A distance cruise control (Adaptive Cruise Control) recognizes the vehicle in front and simultaneously determines its driving speed. This allows it to maintain a desired distance through targeted braking and engine interventions. The active safety systems also include emergency braking assistants. These detect imminent collisions via integrated cameras and environment sensors and, depending on the configuration, initiate partial or full braking. Another driver assistance system is the Electronic Stability Program (ESP), which recognizes imminently dangerous situations such as yawing of the vehicle and then brakes individual wheels in a targeted manner to counteract this rotation. Under certain circumstances, this will prevent the vehicle from swerving.

In addition to comfort and performance criteria, as well as environmental friendliness, passive safety is of great importance in the development of modern and innovative automobile concepts. In particular, crash simulations in the area of occupant protection and partner protection (pedestrians, cyclists or occupants of other vehicles) are essential in an early development phase for the design of the vehicle structure. In doing so, it is important to meet both legal requirements and country-specific consumer protection requirements. This consideration leads to a large spectrum of different calculations and tests. With the help of passive safety systems, the occupants of a vehicle in particular are to be protected from serious or even fatal injuries in the event of an accident. Intelligent systems, consisting of seat belts and airbags, are an important part of passive safety.

In principle, the seat belt is the restraint system that acts first, in the event of an accident. The belt ensures that the occupants are in the intended seating position while driving and that they are restrained in the event of an accident. This prevents vehicle occupants from coming into contact with hard interior parts of the vehicle. If an impact can no longer be avoided in a more severe accident, the seat belt system ensures that the impact speed is reduced. This principle becomes clear in *Figure 1*. It is noticeable here that when the seat belt is worn, the angle α is significantly smaller than is the case without a seat belt. This means that the occupant is braked much more gently by the seat belt and is already being adapted to the vehicle acceleration at time t1, at the start of the locking effect of the belt. This can possibly prevent an impact with parts of the interior and reduce the stress on the occupants. In the case of an unbelted case, the occupant continues to move due to his inertia from the point in time of the crash t0 at a constant speed v until he hits the steering wheel or dashboard at point in time t2 and is severely decelerated.

In order to brake the occupants even more gently, the seat belt must be coordinated with other systems. For example, the seat belt is often designed to be very soft in order to keep chest compression below the critical value, and contact between the body and vehicle parts is then prevented by an airbag. In addition to the seat belt, airbags thus help to avoid impacts with interior parts.

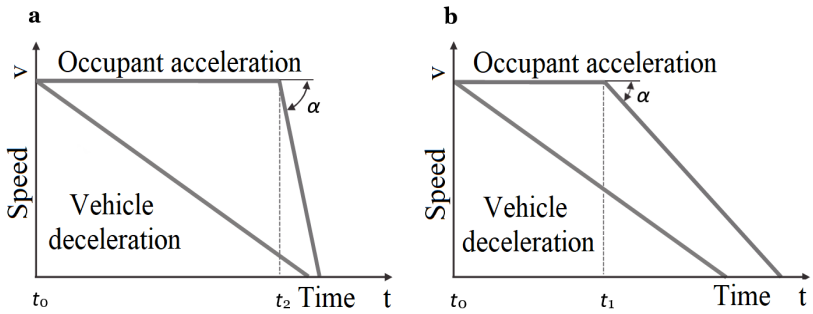


Figure 1
Comparison of the vehicle and occupant accelerations a) without seatbelt and b) with seatbelt [1]

In the event of an accident, the airbag is automatically opened and the occupant is specifically picked up. The idea of this type of restraint arose back in the 1960s. However, airbags were considered problematic and unreliable because of the short inflation time required for the air cushion, the high weight and the large construction volume due to the high-pressure gas cylinders used. The necessary filling time could only be achieved with the development of solid gas generators in the 1970s. Initially, airbags were offered at extra cost in luxury vehicles, whereas nowadays airbags are standard equipment in almost all vehicles.

A large number of sensors detect acceleration and forward these signals to the control unit for processing. When dangerous situations are detected, the gas generator is activated. The ignition pulse, which is generated by the trigger electronics, ignites a pyrotechnic propellant charge. A gas mixture is released and unfolds the airbag. The time required for this is between 30 ms and 40 ms on the driver's side and between 40 ms and 60 ms on the passenger's side. A schematic structure of an airbag system is shown in *Figure 2*. The airbag is in a folded state here.

The energy dissipation of the airbag is determined by the dimensioning of the outflow openings, so-called "vents", and the air permeability of the membrane fabric. The specially designed "vents" ensure that the gas mixture flows out precisely and regulates the pressure in the airbag. This prevents the occupant from being thrown back into the seat backrest. In addition, the pressure regulation serves to ensure that the vehicle occupant gently immerses into the airbag and thus the targeted energy absorption of the occupant kinematics. The outflow openings are always integrated on the back of the air cushion and thus face away from the vehicle occupants. When the relative speed of the vehicle occupants to the body is reduced, the air bags collapse. This happens about 120 ms after the impact.

Due to the larger installation space on the passenger side and the lack of a steering wheel, the passenger airbag has to fill a significantly larger volume than the driver's airbag.

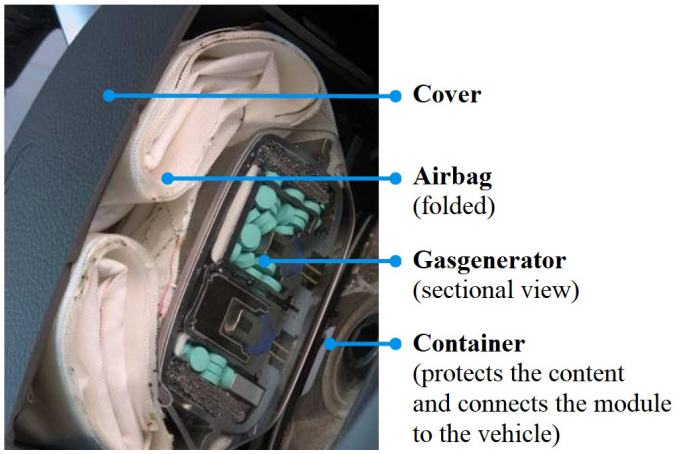


Figure 2
Structure of an airbag module [2]

In addition, the volumes differ due to the different legal situation in the European countries and the United States. Furthermore, the pressures when the occupant enters the airbag are designed differently. The immersion pressure on the driver's side is usually p = 0.6 bar and on the passenger side between p = 0.1 − 0.4 bar. A significant difference can also be observed regarding the size of the airbags, the typical size of the driver airbags is 45-60 liters, while the passenger airbags are much larger, between 80-130 liters. [3]

Not only the size of the airbag, but also the shape of the air cushion is important. The driver's airbag is rotationally symmetrical and circular. It is usually integrated in the steering wheel. This shape enables the driver airbag to behave in the same way at every possible steering angle. In contrast, prismatic, rectangular shapes are preferred for a passenger airbag. Coordination of the head is an important aspect of the design of the airbags on the passenger side. To this end, the “butterfly”, shaped passenger airbags, have appeared in recent years, which thanks to their design, are able to direct the head towards the center of the airbag and reducing the loads on the head in case of oblique collisions. Driver airbags and front passenger airbags, alongside the seat belt, are now state-of-the-art technology from restraint system. Basically, airbags can be designed as single-chamber or multi-chamber systems and also offer a great deal of leeway in the overall design of the passive safety systems.

3 Description of the EuroNCAP Side Crash Load Cases

In order to be able to evaluate the safety of the occupants equally for different scenarios, both frontal and side impacts are taken into account in the EuroNCAP crash tests. There are two frontal load cases, one with a deformable barrier and one with a rigid wall. In the side load cases, a pole and a deformable barrier are used as crash partners for the vehicle. Various loads on the structure and the dummy used are measured. The greatest challenge of all side load cases is that, compared to frontal load cases, there is almost no vehicle structure between the crash partner and the occupant that can already take a lot of energy out of the system through deformation before the impulse reaches the occupant. Consequently, severe injuries often occur in these crash situations. In order to be able to estimate the risk of injury, key figures were introduced over time. [4]

For example, a very important value for assessing the severity of a head injury is the Head Injury Criterion (HIC). The acceleration a(t) is measured here in the center of gravity of the head and is included in the criterion as a multiple of the gravitational acceleration g. The time interval considered [t1, t2] is either 15 ms or 36 ms long, which is why HIC15 and HIC36 are often also written for a more precise distinction. Furthermore, there are also several values for this, since the time intervals considered are usually longer than 15 ms or 36 ms. In this case, the maximum occurring value must be taken. The importance of this criterion can be seen in *Figure 3*, for example an HIC value of 1000 means that there is a 50% probability of irreversible injuries occurring in the occupant's head area.

The Maximum Abbreviated Injury Scale has a scale from 1 to 6, where level 1 is minor (superficial laceration), level 2 is moderate (fractured sternum), level 3 is serious (open fracture of humerus), level 4 is severe (perforated trachea), level 5 is critical (ruptured liver with tissue loss) and level 6 is maximum (total severance of aorta) injurys means.

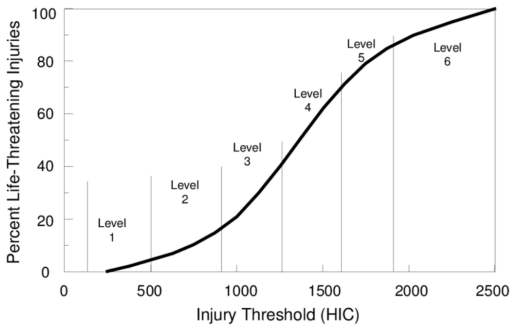


Figure 3
Maximum Abbreviated Injury Scale [5]

Other measurement points in the dummy, such as the compression of the chest or the force on the hips, are also used for the evaluation. Points are then distributed according to a prescribed point system, which then go into an evaluation scheme for the Adult Occupant Protection section of EuroNCAP. This section describes the test execution and evaluation for test cases with adult occupants. Overall, the ratings from this sub-area and other sub-areas, such as test cases for underage inmates, are combined in a star rating. These overall ratings are then published by EuroNCAP.

3.1 Pole Test

The idea behind the pole impact is to simulate a load case in which a vehicle crashes sideways into a tree at low speed. In the test setup, this is a rigid pole with a diameter of 254 mm and a height that is greater than the entire vehicle. The test vehicle is crashed onto this pole at a speed of 32 ± 0.5 km/h under 75° to the longitudinal axis of the vehicle (see *Figure 4*). The target line runs exactly through the center of gravity of the dummy head. For the crash test, the vehicle is positioned on a platform, but in the simulation, this problem can be solved simply by applying the appropriate boundary conditions.

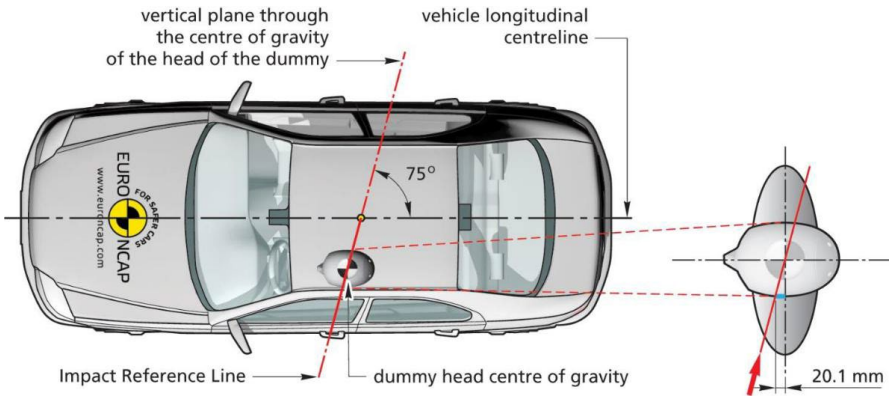


Figure 4
Point of impact of the pole test [6]

In order to be able to estimate the loads and the behavior of the system, numerous simulations are already carried out before the first tests. This is particularly useful because this test setup depicts a situation that is as critical as possible. This is because the accelerations here are lower than in the case of a barrier impact, but the deformation or intrusion is significantly higher. Since the head is close to the action here, this is always more threatening to the occupant.

3.2 Mobile Deformable Barrier Test

Most real side impacts can be classified as being either a wide or a narrow crash partner. The narrow crash partner is to be represented by the pole load case, whereas the wide crash partner is represented by the mobile deformable barrier. The Advanced European Mobile Deformable Barrier used by EuroNCAP is a further development of the original mobile deformable barrier. It consists of a car with four tires and a deformable honeycomb structure at the front. This structure shown in *Figure 5* serves to absorb energy and is intended to represent the front of a larger car. The external dimensions of this structure are around 1700 × 560 mm.

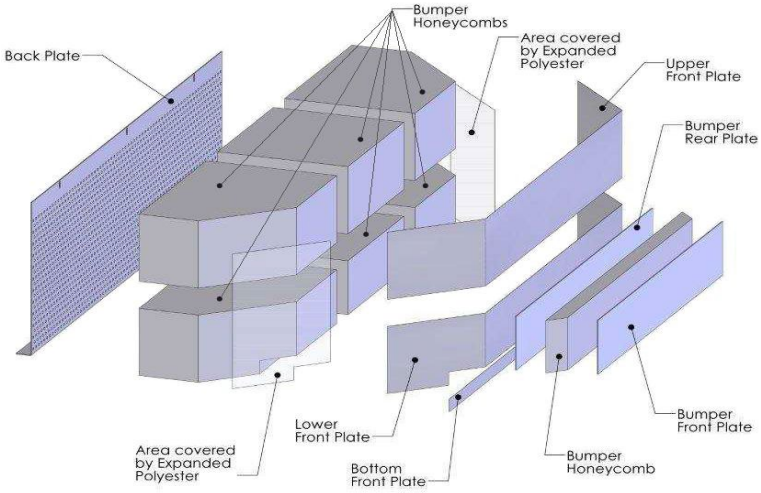


Figure 5
Exploded view of the Advanced European Mobile Deformable Barrier front honeycomb structure [7]

Overall, the barrier should have a mass of 1400 kg. The barrier then hits the vehicle in the area of the two doors at a speed of 60 ± 1 km/h after it has been accelerated to the corresponding speed. The vehicle stands still in this load case. Due to the greater width, less deformation or intrusion occurs in this load case compared to the pole load case, whereas the acceleration is greater due to the greater difference in speed.

In this research, I examined the pole test because this means more stress for the passengers. Car body protection is minimal and deformations are large. During the examination of the side crash, I used the experience and knowledge gained during the examination of the frontal crash.

4 Description of the Simulation System

4.1 Finite Element Method in General

With the help of the FEM (finite element method), complex assemblies can be examined with regard to their static and dynamic behavior. FEM is used in particular in the areas of structural mechanics, heat transfer and fluid mechanics. The basic idea is to divide the structure to be examined into finite elements, which are coupled to each other via nodes. For each of these elements, equations are defined that describe the physical problem. Taking into account the boundary and initial conditions, these can be solved with suitable algorithms. The exact procedure of such a numerical calculation is divided into the three areas of pre-processing, solving and post-processing (see *Figure 6*).

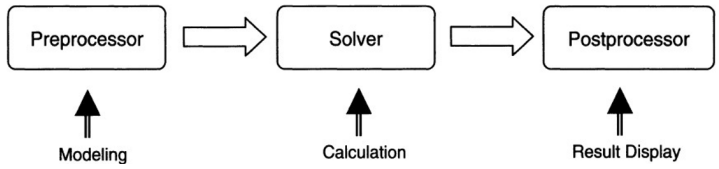


Figure 6
Simulation steps [8]

The model for the respective solver is set up in pre-processing. First, geometry data is imported or generated directly from the preprocessor. Furthermore, components with a finite number of finite elements and their nodes are discretized. The material properties are also defined in this process step and assigned to the respective components. In addition, contacts, boundary conditions and initial conditions must be defined in order to be able to subsequently carry out a model check and transfer the model to the solver. The solver serves as an equation solver and, taking the boundary conditions into account, determines the displacements of the individual nodes, whereupon other variables such as stresses and strains can be calculated for each element with this information. During post-processing, the focus is on evaluating the respective finite element calculation. Here the results are checked with the previously defined evaluation criteria using suitable post-processing tools and presented visually. The results are then documented and the model improved if necessary.

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. [9]

4.2 Basics of Occupant Protection Simulation

In principle, solvers can solve systems of equations implicitly or explicitly. With the implicit time integration, the following equation of motion becomes

$$\underline{\underline{M}}^{t+\Delta t} \cdot \ddot{\underline{u}} + \underline{\underline{C}}^{t+\Delta t} \cdot \dot{\underline{u}} + \underline{\underline{K}}^{t+\Delta t} \cdot \underline{u} = \underline{F}_{ext}^{t+\Delta t}$$

evaluated at the unknown point in time $t + \Delta t$. $\underline{\underline{M}}$ represents the mass matrix, $\underline{\underline{C}}$ the damping matrix, $\underline{\underline{K}}$ the stiffness matrix, $\ddot{u}$ the acceleration vector, $\dot{u}$ the velocity vector, u the displacement vector and $\underline{F}_{ext}$ the acting external forces. With linear relationships of the gradient matrices $(\underline{\underline{M}}, \underline{\underline{C}}, \underline{\underline{K}})$, a solution can be determined by solving the differential equation system once. The calculation of non-linear systems, on the other hand, is carried out incrementally with successive equilibrium iterations. Strongly non-linear structures can lead to convergence problems with implicit time integration, which is why no solution can be found. Therefore, the method of implicit time integration is suitable for linear dynamic or static tasks. [10]

In the case of explicit time integration, on the other hand, the equation of motion is evaluated at the known point in time t:

$$\underline{\underline{M}}^{t} \cdot \ddot{\underline{u}} + \underline{\underline{C}}^{t} \cdot \dot{\underline{u}} + \underline{\underline{K}}^{t} \cdot \underline{u} = \underline{F}_{ext}^{t}$$

Assuming a diagonalized mass matrix and thus the simplification that the masses at the nodes are understood as point masses, the system of equations can be converted into Newton's second axiom and decoupled:

$$\underline{F}_{\text{int}}(\underline{u}, \dot{\underline{u}}, \ddot{\underline{u}}) - \underline{F}_{ext} = 0$$

$$\text{with } \underline{F}_{\text{int}}(\underline{u}, \dot{\underline{u}}, \ddot{\underline{u}}) = \underline{\underline{M}} \cdot \ddot{\underline{u}} + \underline{\underline{C}} \cdot \dot{\underline{u}} + \underline{\underline{K}} \cdot \underline{u}$$

$$\ddot{\underline{u}} = (\underline{F}_{ext} - \underline{F}_{\text{int}}(\underline{u}, \dot{\underline{u}})) \cdot \underline{\underline{M}}^{-1}$$

This means that each equation can be evaluated independently, which is why an overall faster solution is found compared to the implicit method. Since stresses are generated as pressure waves (sound waves) in the material at the speed of sound c:

$$c \approx \sqrt{\frac{E}{\rho}}$$

propagate, it is essential to select the calculation time step in such a way that it is always smaller than the greatest possible information propagation (sound propagation in the component) over the smallest element length. If this stability criterion is met, each discretization point can be calculated separately. Otherwise, there is a loss of information, which means that the solver cannot find a suitable solution using the explicit method.

The critical time step Δt can then be:

$$\Delta t < l_{\min} \sqrt{\frac{\rho}{E}}$$

to calculate. As a result of plastic deformation, it is possible for some elements to become very small. As a result, these very small elements are used to determine the critical time step. This would increase the computing time extremely and lead to calculations that could not be carried out. Therefore, during the explicit calculation, these elements are deleted or the density of these small elements is increased. This artificial scaling of the mass is called "mass scaling" and allows the time step to be increased. [11]

Compared to the implicit calculation, non-linear influences mean only a small additional effort. However, a dynamic calculation must be carried out for static problems. The explicit calculation method is therefore suitable for very short-lasting dynamic processes such as crash, explosion, impact and puncture simulations. Furthermore, explicit solvers can be used for contact problems as well as for calculations of highly discontinuous structures in which no corresponding result can be found with implicit methods.

4.3 Seat Position Definition

In the earlier stage of my research, we examined injuries that occur in frontal crashes. In the event of a frontal crash, we examined a total of five rotated positions, these were 30°, 60°, 90°, 135°, 180° and, of course, the normal non-rotated state. It is absolutely necessary to narrow down the seating positions to be analyzed when examining the effect of swivel seats on passive safety systems. The process of rotating the seat is completely identical to the method used in frontal crashes. Rotating the driver's seat is impossible without modifying the interior of the vehicle, so I took all previous modifications from the frontal crash.

In the case of a frontal crash, it was observed that the maximum value of HIC15 is reached in the simulation model already at a rotation angle of 30°. There was a risk of serious head injuries due to the changed impact point of the head, which is no longer in the center of the airbag, the protective effect of the airbag is significantly reduced. The test head does not optimally reach the center of the airbag for all angle variations. In the 60° version, the steering wheel is hit directly. The seat with an angle of 60° is the worst for the airbag and the highest damage values occur here, so we will examine this case in the case of a side impact.

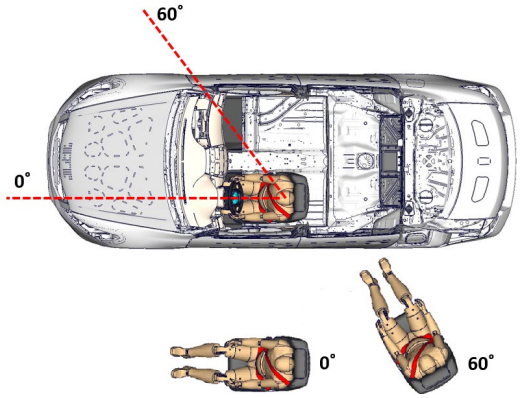


Figure 7
Representation of the examined angles of rotation (Source: Author's plot)

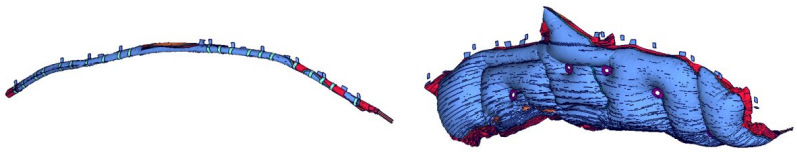
The frontal crash investigation showed that, as a result of the modified seating positions, the driver's movement kinematics changes radically in the event of an accident. We found that the effectiveness of the passive protection systems is greatly reduced and the risk of fatal injuries is greatly increased in the case of inverted seating positions. This is especially true for head and neck injuries, but the value of chest compression and the extent of leg injuries also become more severe. In this research, we examine the effect of this on side crashes.

5 Results of the Simulations

In the following chapter, an overview of the results of all examined angles of rotation during a side pole crash test 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 general, all necessary side crash restraint systems are deployed for the 0° variants. This includes the side airbag and the head airbag. The airbags used here are head or curtain airbags, which are primarily intended to protect the head area

of the occupants of the first and second row of seats. It is installed under the interior from the A-pillar to behind the second row of seats along the entire side and, when it is triggered, unfolds like a protective curtain between the heads of the occupants and the vehicle structure. The side airbag is also intended to prevent the occupant from coming into direct contact with the vehicle structure, but in its position it prevents contact between the trunk area and the door panel or the paneling of the B-pillar. Due to its location in the first row of seats, this airbag is often referred to as a front-side airbag (see *Figure 8*).



Non-deployed curtain airbag Deployed curtain airbag

Non-deployed side airbag Deployed side airbag

Figure 8
Relevant airbags in side impact (Source: Author's plot)

A conflict between the created headrest and the head airbag arises from a rotation angle of 60° or greater, since the effective range of these two restraint system overlaps. From a rotation angle of 60°, the head airbag can no longer unfold correctly and completely and thus loses its protective effect. By rotating the seat, the proportion of the side crash changes with increasing angle in the direction of a rear-end crash during the impact. Since the most important restraint system for a rear-end collision is the headrest, the head airbag is not used for the 60° rotation angle. Therefore, for good comparability, not used the head airbag in the 0° version either. *Table 1* shows the values determined for the angles of rotation 0° and 60°.

It should be noted that there is no center console due to the geometric adjustments to the vehicle model. In a conventional vehicle, the seat is supported by the center console, among other things, in the event of a side impact. This restraint does not exist in the fitted simulation model. The trend of falling head values can be clearly explained with the impact point of the pole in the side structure of the vehicle.

Table 1
Results overview

Criteria	Limit value	Unit	Simulation 0°	Simulation 60°
Head (HIC15)	700	[-]	592	141
Head (a3ms)	80	[g]	77.6	41.7
Head (BrIC)	1.05	[-]	1.01	0.89
Neck (Nij)	0.85	[-]	0.41	0.69
Chest (Compression)	60	[mm]	27.4	37.1
Abdomen (Compression)	88	[mm]	53.7	61.2

The orientation of the pole is adjusted to the 0° seating position and no realignment takes place for the selected seating position for the purpose of comparability. According to this, the point of impact changes from the head with increasing angle degrees. For the THOR dummy in the side crash, the head values in the 60° position are significantly reduced compared to the upright and straight-ahead sitting position. In particular, the HIC values are reduced by almost 80% to 141. In addition to the changed point of impact, the seat has a damping effect. It should also be added that the position of the seat has been pushed further into the interior with a translation. As a result, the distance between the head and the post is too great, so that no serious consequences can be seen.

The turning angle of 60° changes the character of the side crash to a rear crash. This can be seen from the slight hyperextension of the neck, which is evidenced by the increase in the Nij value. There is also an increase in chest indentations with decreasing HIC values, because the protective effect of the seat back does not provide sufficient stability to respond to the intrusions from the pole. It should be added that due to the rotation of the seat, the side airbag integrated into it does not provide sufficient protection for the selected rotating seat scenarios due to the lack of support. For illustration, the maximum intrusions of the pile at the rotation angles of 0° and 60° for the THOR dummy are shown in *Figure 9*.

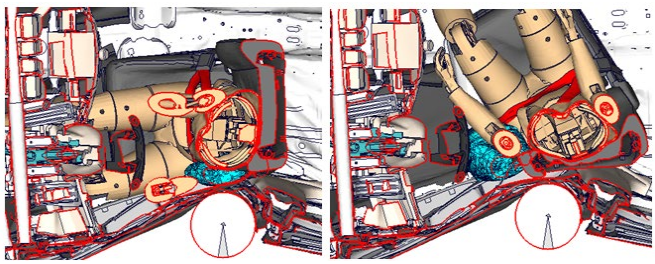


Figure 9
Pole impact on the THOR dummy 0°(left) and 60°(right) (Source: Author's plot)

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. However, before validation a first way of checking the plausibility of the calculations is to look at the crash simulation visually. This can be found for the pole side crash for selected points in time in *Figure 10*. This visual inspection should be done together with an inspection of the energy flows.

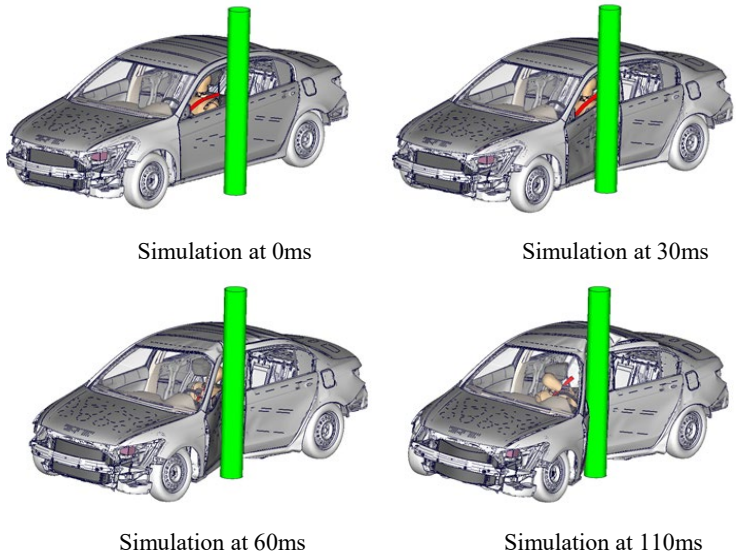


Figure 10
Selected points in time during a side pole crash (Source: Author's plot)

During an inspection of the energy flows a distinction is made between three different types of energy. The kinetic energy of the system, the internal energy and third the hourglass energy. One can see very well that at the beginning the total energy consists purely of the kinetic energy. After that, the kinetic energy drops to a plateau and the internal energy also increases to an almost constant value up to about 10 ms. This is the case because the engine block is very heavy and deflects laterally, which is also the reason for the fluctuations in the total energy at the beginning. The initial contact with the stake happens just before 10 ms. From this point on, the kinetic energy of the vehicle is gradually converted into internal energy through plastic deformation until the vehicle is almost stationary after around 110 ms. Since under-integrated elements are used here, the hourglass energy with which the elements are provided increases over time in order to prevent the hourglassing effect (see *Figure 11*).

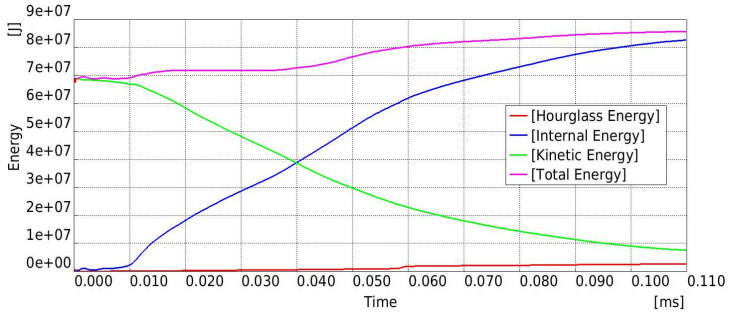


Figure 11
Course of the energy at the pole side crash (Source: Author's plot)

The tested model of the Honda Accord introduced in 2008 was released in Australia. It also included dual front airbags, side airbags and head protection airbags as part of the standard equipment. In terms of active safety, anti-lock brakes system (ABS), electronic brake distribution (EBD) and electronic stability control (ESC) were also standard equipment. In addition to all of this, an intelligent seat belt reminder is installed on every seat. In the front row, the seat belt buckles are mounted on the seats and the upper anchorage points are adjustable. These features greatly improve the effectiveness of the seat belt, thereby increasing safety. Belt tensioners were installed on the seat belts in the first row of seats, so that in the event of a collision, they are able to reduce the slack in the belt. The middle rear seat is equipped with a three-point seat belt. This can provide better protection than a conventional two-point seat belt. [12]

The Accord scored the maximum 8 out of 8 in the side pole crash test (5 Stars). The passenger compartment held its shape well. Airbags and seatbelts tailor the timing and the degree of restraint to suit the size of the occupant and the severity of the impact. In this case, both the driver and passenger were well protected. The Honda Accord scored maximum points in the car side impact test. In the pole side impact, the chest was adequately protected and there was good protection of all other body regions. The head restraint provided good protection against whiplash injuries.

The car includes a passenger airbag deactivation function so that the rear-facing child seat can be used safely in this position. However, the driver does not have enough information about the current state of the airbag. In any case, a warning label clearly warns of the danger of using a rear-facing child seat in the passenger seat without first deactivating the airbag. Unfortunately, this information is not available in all European languages. The existence and exact location of the ISOFIX points on the rear outer seats is not clearly marked, which can also cause confusion.

Figure 12
Honda accord 2008 EUNCAP side pole crash test [12]

Table 2
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	541	575 (+6%)	592 (+9%)
Head (a3ms)	[g]	80	71.4	72.9 (+2%)	77.6 (+8%)
Head (BrIC)	[-]	1.05	0.86	0.93 (+8%)	1.01 (+17%)
Neck (Nij)	[-]	0.85	0.62	0.44 (-29%)	0.41 (-34%)
Chest (Compression)	[mm]	60	25.8	27.0(+5%)	27.4 (+6%)
Abdomen (Compression)	[mm]	88	58.9	55.2 (-6%)	53.7 (-9%)

Table 2 shows the results of the basic model, the version without modifications, and the modified model compared to the results of the crash test. The basic model is validated, it can be seen that all deviations in the simulation results remain within the appropriate limits. The situation is similar with the modified model, because the modifications mainly affect the frontal collision only. The only major difference is in the values of the neck, the main reason for this being the replacement of the seat belt attachment point. The modified model must be used for comparison with the rotated seat versions, as these modifications allow the driver's seat to be rotated. [13]

Conclusions

The aim of this research was to investigate a pole side crash, in the case of a fully self-driving vehicle. In order to do this, we used as a basis the computer simulation model used in the frontal collision investigation. We have created the necessary definitions of side impact from the point of view of passive passenger safety. We validated the completed computer model with a real crash test and then transformed it so that it is suitable for testing the rotated seating positions. In the case of a side impact, We examined the most critical rotation angle in the case of a frontal impact. We examined to what extent the seating position of the passengers affects their injuries in the event of a side pole collision. During the simulations and the evaluation, we compared the results of the injuries that occurred in the

case of seat positions turned by 60° with the case of the driver's seat in the normal basic position. As a result, we received that although the kinematics of the driver's movement changes radically during an accident due to the modified seating positions, the degree of injuries is still greatly reduced. Exceptions to this are injuries to the chest and abdomen area. This decrease in injuries can be explained primarily by the damping effect of the driver's seat, because in the case of the driver's seat turned by 60°, the driver's seat protects the passenger from the direct load and deformation coming from that side. Based on the results, although the damage values only partially increase in the case of side crash, it is still necessary to expand the passive safety system. Next goal of the research is the creation of a protection system capable of providing protection even in a seat position rotated at any angle. This system must be able to provide sufficient protection in both frontal and side collisions. In order to achieve this goal, the possibilities of further developments for the driver's seat need examination. The seat must be able to coordinate the movement of passengers in the event of any accident and must participate in the prevention of serious injuries and the absorption of large deformations a model to test and verify this will be built. In addition to the possibilities for further development of the driver's seat, examination is needed for future new possibilities of the use of airbags, in self-driving vehicles. The future development of the airbags as known today, will be inevitable, in a vehicle without a steering wheel. The aim of the research will be to define the challenges that affect airbags and to define any possible solutions. It will certainly be necessary to examine the new position, shape and size of the airbags known today, in order to achieve the desired goal.

References

[1] Kramer F., “Integral Safety of Motor Vehicles: Biomechanics - Simulation - Safety in the Development Process”, Springer Vieweg, Wiesbaden, 2013

[2] Schidler V., Kühn M. „Intelligent restraint systems“, Reports of the Federal Highway Research Institute, Bremerhaven, 2004

[3] C. R. Bass, J. R. Crandall, J. R. Bolton, W. D. Pilkey, N. Khaewpong, and E. Sun, “Deployment of air bags into the thorax of an out-of-position dummy”, SAE Technical Paper, paper no. 1999-01-0764, pp. 1-17, 1999

[4] S. Umale, M. Arun, H. Hauschild “Quantitative evaluation of THOR world SID and hybrid III under farside impacts”, Conference proceedings International Research Council on the Biomechanics of Injury, 2018

[5] 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

[6] European New Car Assesment Programme “Oblique pole side impact testing protocol”, 2019

[7] European New Car Assesment Programme “Technical Bulletin: AE-MDB Specifiation”, 2013

[8] P. Fröhlich, “FEM application practice. Introduction to Finite Element Analysis”, bilingual edition german/english, Studium Technik, 2005

[9] J. O. Hallquist, LS-DYNA R Theory MManual, Livermore Software Technology Corporation, Livermore, California, 2006

[10] Zienkiewicz O. C., Taylor R. L., ZHU J. Z. “The Finite Element Method: Its Basis and Fundamentals” 7th edition, Elsevier Science, Burlington, 2013

[11] Desai C. S., Abel J. F. “Introduction to the finite element method: A numerical method for engineering analysis”, Van Nostrand Reinhold, New York, 1971

[12] ANCAP, “Honda Accord 2008 crash testing for safety”, New Car Assessment Program, Australia, 2009

[13] H. Wu, H. Hou, M.Shen “Occupant kinematics and biomechanics during frontal collision in autonomous vehiclescan rotatable seat provides additional protection”, Computer Methods in Biomechanics and Biomedical Engineering, 2020

Energy Losses during Intestine Deformation – A New Method of Robot Propulsion?

Łukasz Frącczak, Paweł Żak, Michał Starosta

Institute of Machine Tools and Production Engineering, Lodz University of Technology, Stefanowskiego 1/15, 90-924 Lodz, Poland;
lukasz.fracczak@p.lodz.pl, pawel.zak@p.lodz.pl, michal.starosta@dokt.p.lodz.pl

Abstract: Currently, numerous studies are conducted on diagnostic devices for the digestive system. To correctly design such a device, it is necessary to fully specify the parameters of the environment in which the device is supposed to operate. Most of the research conducted on the intestines focuses on mechanical and tribological parameters connected to work with diagnostic capsules. Yet, to simulate the work of the robot moving with a snake-like motion it is necessary to estimate the level of energy dissipated by the intestine deformation as it serves as a main driving factor, which is proven in the paper.

Keywords: intestine mechanical parameters; energy dissipation; robot propulsion

1 Introduction

The development in the field of mobile robotics is determined by the knowledge of the work environment. The same goes for robots prepared to move inside the human body. One good example of such a case is in the field of robotics focused on the diagnosis and treatment of the human digestive system. This field is extremely important, as it aids doctors in the treatment of digestive system disease, including intestinal cancer, therefore, many studies are being conducted on developing the most universal tool. The most commonly used method is the capsule endoscopy [1], which utilises a small-volume capsule that travels through the digestive system, taking photos or picking tissue samples in the process. Yet, there is hardly any control over such capsule motion. An interesting example of a different solution is a robot able to move inside the intestines using a snake-like motion [2]. In Figure 1 an actual model of a named robot is presented. The given picture shows the robot's motion capabilities. It is worth mentioning that because of its construction, it is able to generate sine motion, which is necessary to make its motion possible in any given position or configuration of the body. The dimensions of the presented model currently are: 15 mm of diameter and ca. 1.5 m of length. The remaining question is: will it be able to move inside the human body according to the assumptions?

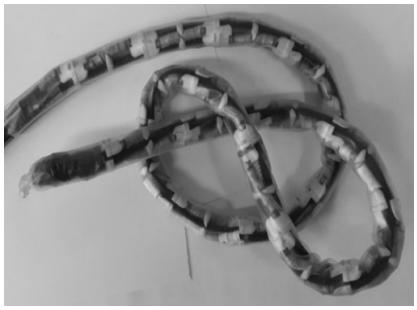


Figure 1
Actual model of the robot moving with snake-like motion

This type of motion is difficult in this environment, as it acts like a visco-elastic body [3]. Assume that a robot moving using a snake-like motion (close to sine) would deform the intestines in a direction perpendicular to its own body (Fig. 2).

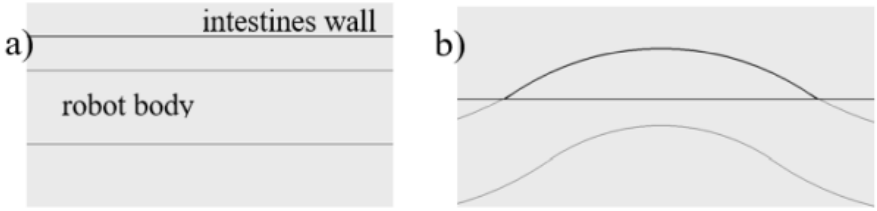


Figure 2
Deformation of the intestine by the robot: a) before deformation, b) after deformation

In this case, a 2D static model of the robot–intestine system shows reaction forces between the robot and the intestines, without internal forces analysis, will take the form shown in Fig. 3. The additional assumption made was that the robot's deformation will take a shape close to a circle.

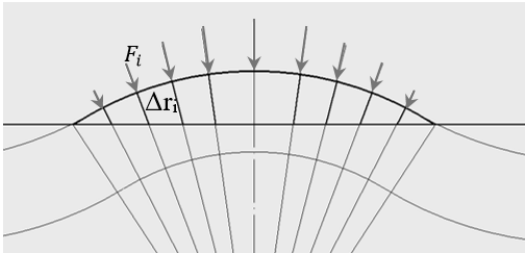


Figure 3
Reaction forces between the robot and the intestines

For this assumption, the work needed to be done to deform the intestines can be given by the following equation:

$$W = E_p + \Delta E \tag{1}$$

Where W – work done to deform the intestines, E_p – potential energy accumulated in the deformed intestines, ΔE – energy that was dissipated in the intestines and cannot be reclaimed.

With the assumption that the intestines are being deformed perpendicularly to the robot's surface, the deformation energy of the intestines can be given as:

$$E_p = \frac{1}{2}\sum_{i=1}^{n} F_i * \Delta r_i \tag{2}$$

Where: F_i is the force with which the robot deformed the intestines in the i-th contact point, Δr_i – the intestine deformation in a direction normal to the robot's surface.

Equation (2) is true for symmetrical deformation, i.e., for the robot moving perpendicularly to the intestinal wall. In the case of the robot generating a propulsion wave along its body, the part of the intestines in front of the wave is being deformed while the part located after the deformation wave reacts to the robot with the force generated by its elasticity. In this case, the geometry of the system can be presented as in Fig. 4.

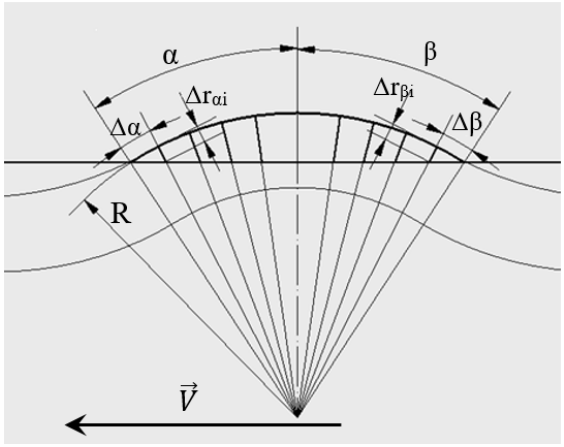


Figure 4
The distribution of deformation geometry for the robot moving with a snake-like motion

The equations of deformation energy can be defined as:

$$W_L = \frac{1}{2}\sum_{i=1}^{n} F_{\alpha i} * \Delta r_{\alpha i} + \Delta E_L \tag{3}$$

$$W_R = \frac{1}{2}\sum_{i=1}^{n} F_{\beta j} * \Delta r_{\beta j} + \Delta E_R \tag{4}$$

Where: W_L, is the work that need to be done to deform the intestines, W_R is the work done by the intestines regaining their original shape, $F_{\alpha i}$ – is the i-th force acting on the left side of the robot, $F_{\beta i}$ – is the j-th force acting on the right side of the robot, $\Delta r_{\alpha i}$, $\Delta r_{\beta j}$ – are the i-th and j-th deformation of the intestines done by the robot at the left and right side. $\Delta E_L = f(\alpha, D, T, t)$ and $\Delta E_R = f(\beta, D, T, t)$ are the energies dissipated during the deformation during the elongation and shape regaining phases of the intestines.

In case of respectively small increases of $\Delta r_{\alpha i}$ and $\Delta r_{\beta j}$ equations (3, 4) take the following form:

$$W_L = \frac{1}{2}\int_0^{\alpha} F(\alpha) * r(\alpha) d\alpha + \Delta E_L \tag{5}$$

$$W_R = \frac{1}{2}\int_0^{\beta} F(\beta) * r(\beta) d\beta + \Delta E_R \tag{6}$$

For symmetrical deformations (case of the robot moving perpendicularly to intestines wall) and for perfectly elastic deformations (ΔE_L=0 and ΔE_R=0), it can be stated that $W_L = W_R$. Yet, in the case when the robot moves with a snake-like motion, the work done by the intestines regaining their shape is equal to the potential energy accumulated in the intestines and can be given with equation:

$$E_p = W_L - \Delta E_L = W_R \tag{7}$$

By including equations (5), (6) in (7) the following is obtained:

$$\frac{1}{2}\int_0^{\alpha} F(\alpha) * r(\alpha) d\alpha = \frac{1}{2}\int_0^{\beta} F(\beta) * r(\beta) d\beta + \Delta E_R \tag{8}$$

It must be stated that the total deformations on the left and right sides of equation (8) must be equal. Otherwise, the intestines would remain permanently deformed, which is unacceptable. Therefore, the following equation is true:

$$\int_0^{\alpha} r(\alpha) d\alpha = \int_0^{\beta} r(\beta) d\beta \tag{9}$$

By comparing equations (8) and (9), it can be seen that energy loss is bound with the reaction forces that caused the deformation. By assuming that the system

shown in Fig. 2 is static and by projecting all the forces onto the direction of the wave, the following is obtained:

$$F_L = \sum_{i=1}^{n} F_{\alpha i} * cos\alpha_i \tag{10}$$

$$F_R = \sum_{i=1}^{n} F_{\beta j} * cos\beta_j \tag{11}$$

and from (7) and (8) it is known that $F_L > F_R$

It can be stated that:

$$F_c - T = F_L - F_R \tag{12}$$

Where: F_c is the total force of intestinal reaction on the robot in the direction of wave propagation, T – is the friction force, F_L, F_R are the forces respectively before and after the deformation wave.

Equations (12) and (8) show that the forces needed to be used to deform the intestines are greater than forces generated by the intestines regaining their original shape, therefore by generating the snake-like motion force Fc is generated. It is directed along with the robot and reverses to the direction of the wave of displacements. Equation (8) clearly shows that propulsion forces will increase proportionally to the volume of Energy dissipated during intestine deformation. It shows the necessity of checking how much of the energy is dissipated during intestine deformation. The proposed research will provide data necessary to estimate the volume of force possible to be generated by a robot moving with a snake-like motion.

Increasing understanding of the mechanical characteristics of the intestines took on particular importance in the context of the development of new medical instruments (e.g., capsule robot, intellectualized endoscopy). A significant element of this understanding is gaining information about the mechanical influence between the intestinal tissue and the instrument.

To obtain information about the stresses and strains (and their relations) which accompany the biomechanical changes during functional loading and unloading of the human gastrointestinal (GI) tract, a reference was made to Hans Gregersen's research [4]. As a basis for 3D anatomical models, digital images gained through ultrasound, using computer tomography (CT) or magnetic resonance (MRI) were used. Models were analyzed using different mathematical algorithms - the finite elements method (in this particular example of the mucosal folded three-layered esophagus) and assuming that it is a thin-walled isotropic structure (e.g. stomach antrum wall surface model). The finite element method provides the possibility to overcome the major shortcoming connected with the thin wall assumption.

The research, focused on gaining information about the mechanical characteristics of the intestines, included two aspects of influence between the medical instrument and the intestinal tissue: the "frictional force" and the biomechanics of GI tract tissue.

The rest of the paper is organised as follows: Sections 2 and 3 focus on state-of-the-art clarification, while Sections 4-6 describe tests done on actual tissues including a description of a tests stand, used methods and drawn conclusions.

2 Frictional Force

The frictional resistance force includes the nominal frictional force [5] and the visco-adhesive force and it depends on the size, material, mass, shape, contact surface contour, and the velocity of the potential instrument (e.g. capsule robot) [6] [7]. The nominal frictional force is related to the elastic restoring force, the real frictional force, and the contact angle; the cohesive force is determined by the contact area and the contact angle [7]. Ex-vivo experiments were carried out in the porcine small intestine with fifteen capsule robots (capsule endoscope) with different shapes and dimensions [8]. The friction increased accordingly with the speed and capsule dimensions, especially by the dimensions perpendicular to the moving direction. To predict the behavior of the capsule robot in clinical tests, the analytical friction model of the capsule robot in the small intestine has been researched [9]. In addition, the results of a similar experimental investigation proved that the Coefficient of Friction (COF) did not change significantly concerning the apparent area of contact between the capsule and the intestine. The COF decreased with an increase in the normal load and varied from 0.08 to 0.2 [10]. In the case of capsule robots with their locomotive mechanism, the stroke of the capsule also influences the friction force [11]. During consideration of the self-propelled robotic endoscope concept, it is extremely important to keep in mind that locomotion in the small intestine is further complicated by the fact that the intestine is not a rigid pipe and is susceptible to damage [12]. The presented analysis states that the friction coefficient between the intestines and the endoscope capsule or robot takes an important role in moving the device inside the intestine, but the research has been conducted only on the motion of the endoscope capsule. In the case of the drive of a device moving with snake-like motion, the key will be the mechanical parameters of the intestines in addition to friction forces.

3 Intestines Mechanical Examination

An example of mechanical tests of research on the viscoelasticity of intestines was obtained by the use of a Dynamical Mechanical Analyzer (DMA) [4]. In this work, it was proven that the dynamic mechanical parameters of the intestinal tissues are not always steady, especially when forced oscillation happens. In practice, this means that the movement of the medical instrument would work in a changing environment, and during the modeling of the intestine, the variation should be taken into account. Based on the sheer measurement of the DMA and the stress relaxation measurement, a five-element model describing the viscoelasticity of intestines was obtained. Additionally, the DMA tests indicate that the storage modulus descends with the increasing shear strain, which means the hyper-elasticity of the small intestine can be measured numerically. As a consequence, the obtained data could be the basis for numerical calculations and the relaxation modulus could be also calculated [13].

Similar experimental investigations of stress relaxation and the stress-strain relations have also proven that the small intestine shows the typical behavior of a viscoelastic material [14]. Dynamic testing results show that the storage modulus of the intestine decreases first and then increases when the frequency is raised within the range of 1-20 Hz [14].

Results of an experimental study of the in vivo and ex vivo compression of goat large intestine revealed a significant difference between the results at the lower and higher rates in in-vivo tests (at the lower rates, the tissue appeared softer) [6]. What is also important, there was a difference between the results of the in vivo and in vitro tests. In vitro results showed less dependence on the compression rates, when the in vivo results were strongly dependent on the compression rates. There is a risk of using only in vitro data in numerical simulations of modeling tissue [6]. An analogous experimental study of the in vitro compression of dog intestine presents very comparable results. The shape and parameters of the curves obtained were very close to the analysis of goat tissues. The incremental elastic modulus of the duodenum of the dog was larger than that of the jejunum under particular pressure [15].

The tensile properties of the human GI tract tissues (including the small and large intestine) generated for cadaveric and surgically removed specimens have been also examined in several experimental investigations. The values of maximal stress and destructive strain were the following: for small intestine transversal specimens - 0.9 MPa and 140% and for large intestine transversal specimens - 0.9 MPa and 180%. [16] The submucosa and muscular layers condition the mechanical strength of the intestine wall; serosa and mucosa showed no significant strength.

As shown, tests on the intestines' mechanical parameters are nothing new. Most of the research has been conducted focusing on selected device types, mainly a

Capsule Robot (CR). Yet, to date, no research had been done to test the intestines with a snake-like motion. It will be crucial to test viscoelasticity [4] and damping [13] factors, friction forces and to estimate the energy dissipated in the intestines. The parameters mentioned at the end are definitive for the use of the phenomena that while deforming the intestines the force applied to deform them is greater than the force generated by intestines returning to their original shape. The difference in forces can be used to propel the device moving inside the digestive tract. The interaction between selected parts of the intestines will be important along with the energy dissipated during the intestine elongation. Another important task connected to the determination of the intestines' mechanical parameters is the possibility to implement them into simulation software. The more accurate the parameters are, the more accurate the obtained simulation results will be.

4 Materials and Methods

Tests on the dynamic parameters of the intestines were made using the custom-made test stand presented in Figure 5. It consists of a linear drive with a positioning accuracy of 0.01 mm and regulated feed in the range from 0 to 100 mm/s. On the carrier of the drive, a 6-axis force sensor has been placed along with a diagnostic probe.

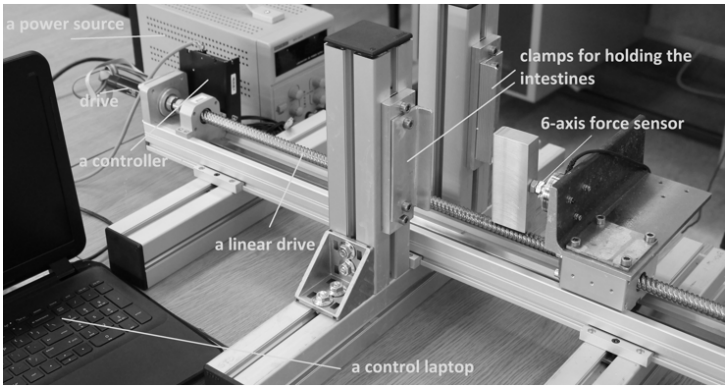


Figure 5
A test stand

The stand has been used to perform numerous tests the purpose of which was to estimate the intestines' elasticity curve and to estimate the volume of the energy dissipated during their deformation. The intestines were taken from pigs just after being slaughtered. During transport, they were kept at a temperature of 3° C, next they were initially cleaned and placed in saline. Before the experiment, the

samples were heated to the ambient temperature of ca. 23° C. The time between taking the sample and the test ranged from 40 minutes for the first sample up to 90 minutes for the last one. The tests were made with forces not causing damage to the samples. During the preliminary test phase, ten repeats were done to specify minimal breaking force, which was estimated at ca. 20 N. Therefore, the tests were made keeping the forces under 15 N.

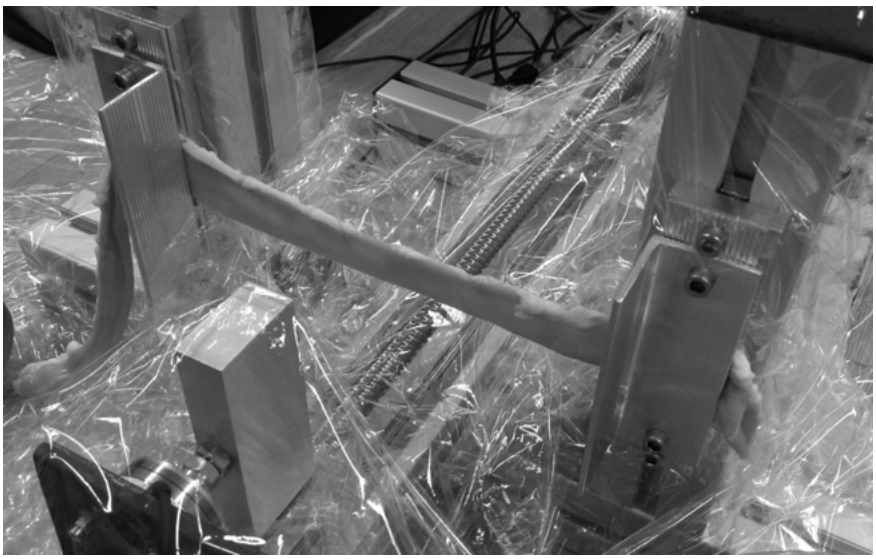


Figure 6
A fixed intestines sample

The sample was placed in holders (the distance between them was 87.3 mm) on the stand presented in Figure 6. The intestines were initially stretched with a force of ca. 3 N magnitude. Next, the diagnostic probe was moved toward the intestines while keeping the force sensor level of 0 N. After that, the test probe was moved in a direction perpendicular to the intestinal axis stretching them, and then it was moved back to the initial position. The displacements and forces generated were recalculated into displacements in a longitudinal direction. Such displacements were made five times for five simples with three different elongations each time. The results of the sample tests can be seen in Figure 7. It shows that the forces of stretching the intestines (right direction) are greater than the forces of the intestine acting on the sensor during the release motion (left direction). The volume of force dissipated in this system is the difference of the surface areas under the diagram of stretching and releasing, therefore it can be expressed with the following equation:

$$\Delta E = \int_{l_0}^{l} f(l)dl - \int_{l_0}^{l} f'(l)dl = \int_{l_0}^{l} (f(l) - f'(l))dl \tag{13}$$

where: f(l), f'(l) are the stretching and release force curves, l – the maximum length of the intestine in maximum elongation for which motion direction change occurred, l_0 - preliminary elongation of the sample.

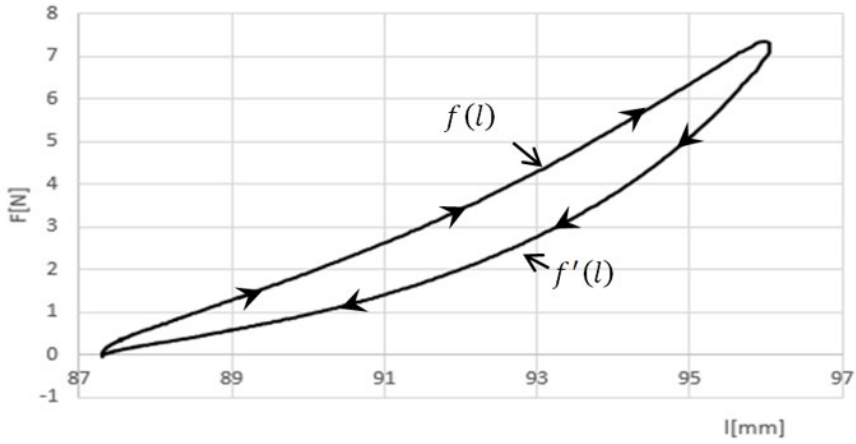


Figure 7
Force curve for longitudal elongation of the intestines by 9 mm (10%) and return to the preliminary position

5 Results

To check how much energy is dissipated in intestines deformations. There were 18 intestines samples tested. The samples were elongated with different values: 7; 10; 13% of their length. Three other samples were tested with different elongation with changing elongation speed from 3.3 to 83.3 mm/s.

The calculation results are summarized in Tables 1 and 2. Table 1 shows the results of five samples and five elongations of the intestines and Table 2 shows the average values of five elongations for each of the five samples. In addition, three different test series with variating elongation values are shown in this table.

Table 1
Value of dissipated energy and maximum forces in intestines elongation by 9mm (10%)

			Stretching					
			1	**2**	**3**	**4**	**5**	**Average**
Sample 1	**Max force**	[N]	7.3	6.8	6.5	6.3	6.1	6.6
	ΔE	[J]	16.8	8.3	8.4	7.0	6.6	9.4
Sample 2	**Max force**	[N]	7.1	6.6	6.3	6.1	6.0	6.4
	ΔE	[J]	16.2	7.7	6.8	7.3	5.7	8.7

Sample 3	**Max force**	[N]	7.9	7.2	6.9	6.6	6.5	7.0
	ΔE	[J]	17.5	8.9	8.0	7.6	6.0	9.6
Sample 4	**Max force**	[N]	7.8	7.1	6.8	6.6	6.4	7.0
	ΔE	[J]	17.2	9.0	6.1	6.0	6.9	9.0
Sample 5	**Max force**	[N]	7.6	7.0	6.7	6.5	6.3	6.8
	ΔE	[J]	16.6	6.8	7.3	5.8	5.3	8.3
Average	**Max force**	[N]	7.6	7.0	6.6	6.4	6.3	6.8
	ΔE	[J]	16.8	8.1	7.3	6.7	6.1	9.0

Table 2

Mean values of energy dissipation and maximum force for varying intestine elongation

3	**7 mm elongation (7%)**						
	stretching						
		1	**2**	**3**	**4**	**5**	**average**
Max force	[N]	4.8	4.5	4.3	4.2	4.1	4.4
ΔE	[J]	10.1	6.0	5.0	4.8	4.4	6.1
	9 mm elongation (10%)						
		1	**2**	**3**	**4**	**5**	**average**
Max force	[N]	7.6	7.0	6.6	6.4	6.3	6.8
ΔE	[J]	16.9	8.1	7.3	6.75	6.1	9.0
	11 mm elongation (13%)						
		1	**2**	**3**	**4**	**5**	**average**
Max force	[N]	10.9	10.0	9.5	9.1	8.9	9.7
ΔE	[J]	28.2	11.5	10.1	9.6	8.8	13.6

Table 3 shows the percentage decrease of maximum forces and energy dissipated for the first and last stretching of the intestine. Figure 7 shows a sample diagram of forces and elongation percentage for the first and fifth longitude elongation.

The next phase of the research focused on intestines' behavior under the greater velocity of stretching and releasing. To do so, the sample was stretched at various velocities. The test started with a stretching (and releasing) velocity of 3.3 mm/s and then it was increased by 8.3 mm/s until it reached a value of 83.3 mm/s.

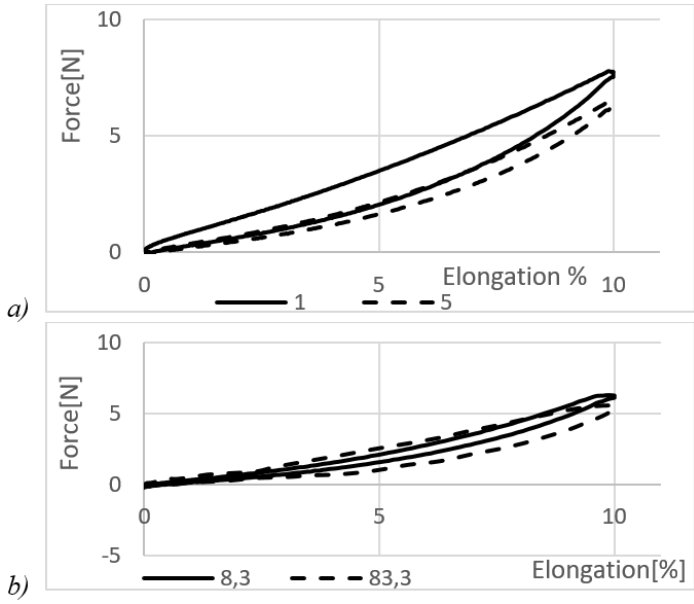


Figure 8
Diagram of intestine elongation by 10%: a) for elongation no. 1 and 5 with 3.3 mm/s velocity; b) with 8.3 mm/s and 83.3 mm/s velocity

Table 3
Mean values of energy dissipation and maximum force for varying intestine elongation

		7 mm elongation (7%)		
		Elongation no.		**Difference**
		1	**5**	**[%]**
Max force	[N]	4.8	4.1	14.6
ΔE	[J]	10.1	4.4	56.4
		9 mm elongation (10%)		
		Elongation no.		**Difference**
		1	**5**	**[%]**
Max force	[N]	7.6	6.3	17.1
ΔE	[J]	16.9	6.1	63.9
		11 mm elongation (13%)		
		Elongation no.		**Difference**
		1	**5**	**[%]**
Max force	[N]	10.9	8.9	18.3
ΔE	[J]	28.2	8.8	68.8

6 Discussion

The obtained results clearly show that the maximum forces of intestinal stretching are greatest during the first stretching and get lower with successive elongations – the lowest values are measured during the fifth elongation. Similar results were obtained for the total energy dissipated in the intestines - the greatest energy value is being lost during the first stretching of each sample and for the longest elongation.

Table 3 and Figure 8a show that for a series of elongations, stretching forces and dissipated energy get smaller. In addition, the results in Table 3 show that for longer elongation, the decrease of dissipation and maximum forces get larger. The presented results indicate that for dynamic interactions, damping properties of the intestines will cause a decrease in the force necessary for another elongation. This can be explained by the decreasing elasticity of samples taken post-mortem, yet this needs to be confirmed during in-vivo experiments.

The next phase of the research focused on intestines' behavior under the greater velocity of stretching and releasing. To do so, the sample was stretched at various velocities. For each velocity, two elongations of the intestines were made and measured values (max force and ΔE) were averaged. Each time between the velocity change there was a break lasting ca. 10 s, the purpose of which was to let the intestines regain their original shape. This interval was selected based on research presented in [14], in which the forces in the intestines stabilized after ca. 10 s (after the elongation). The results of this experiment are presented in Figure 9.

An example diagram for various velocities is presented in Figure 9b. It shows that for increasing velocity, the maximum force drops. The decrease of force in the subsequent repeats of elongations was analogical as in previous tests (Tables 1-3 and Figure 9a).

This phenomenon can be explained by the low ability of samples taken post-mortem to regain their natural shape. This comes with permanent deformation of the samples without causing any actual damage to them. Yet, as a result of energy dissipation increasing with the increase in velocity, it can be stated that these permanent deformations are made at a minimal scale. These are hypothetical theses that need to be confirmed during in-vivo tests. Additionally, the tests show that energy dissipation is greater for faster intestine elongation. This phenomenon can be used to increase the effectiveness of device motion. In summary, the dissipated energy is greater for bigger intestinal elongation and for faster elongation, which is presented in Figure 10.

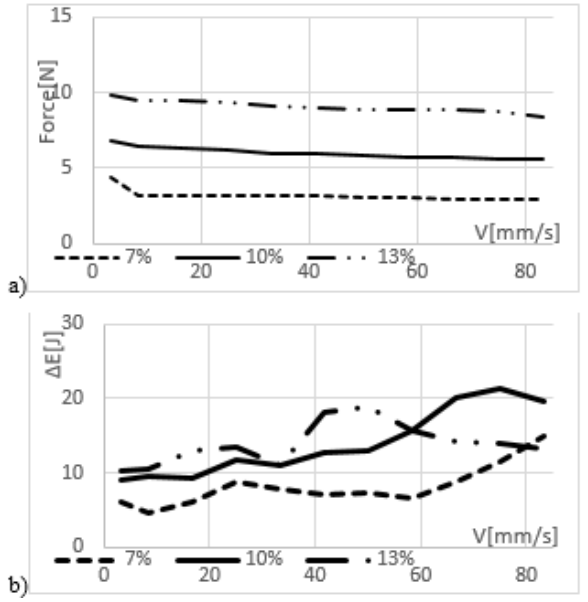


Figure 9
The results of intestine elongation for variety of velocities: a) maximum forces, b) value of dissipated energy

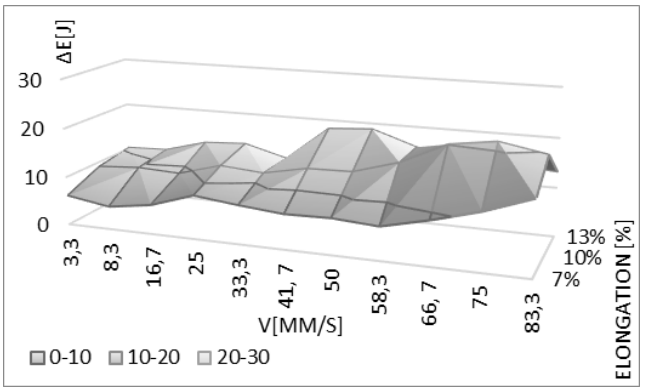


Figure 10
Energy dissipated in the intestines in function of elongation and elongation velocity

Conclusions

The performed experiments show that a series of intestines elongations cause the decrease of forces necessary for subsequent elongations and the decrease of dissipated energy. Tests show that this phenomenon is repeatable, yet it is impossible to state what causes these changes without performing in-vivo tests.

Nevertheless, results show that it is possible to use dissipated energy as the snake-like robot propulsion method.

Another important conclusion is that during the intestine stretching and release, part of the energy is dissipated. Tests show that its value ranges from 4 to 21 J depending on the length and velocity of elongation. It is still necessary to check if the intestines can regain their natural shape during tests with a high velocity between the series. If so, it will result in a lack of energy dissipation. This feature can be used to propel a device that could move inside the intestines using a snake-like motion. The phenomenon of energy dissipation by the intestines during their elongation can hypothetically be a drive for a device moving inside them.

Future works will be focused on creation of the next versions of the robot. It is planned to further increase its length and while decreasing the diameter. Obtained tissues deformation results are to be used as an entry point to perform additional simulations of robot motion inside the intestines. The obtained results of named actions will be the content of future publications.

Acknowledgement

The presented research was co-financed by The National Centre for Research and Development, under LIDER Programme, for 2017-2019, LIDER/20/0106/L-7/15/NCBR/2016

References

[1] Fracczak L, Kobierska A, Koter K, Żak P, Czkwianianc E, Kolejwa M, Nowak A, Socha – Banasiak A, Ślęzak J, The Diagnostic Gastroenterology Needs in Relation to Exisiting Tools, Research and Design Work on a New Tool in Endoscopy Field, Methods and Models in Automation and Robotics (MMAR), 2017 19th International Conference On, pp. 705-710

[2] Fracczak L, Olejniczak M, Podsądkowski L, Long-range snake-like robot powered by pneumatic McKibben muscles, Archive of Mechanical Engineering, Vol. 66, Nr. 2, pp. 257-267, 2019, DOI: 10.24425/ame.2019.128447

[3] L. Fracczak, B. Bryl-Nagórska, P. Żak, A simulation experiment of snake-like robot module, 2018 18th International Conference on Mechatronics - Mechatronika (ME), Brno, 2018, pp. 392-396

[4] Liao D, Frokjaer JB, Yang J, Zhao J, Drewes AM, Gilija OH, Gregersen H. Three-dimensional surface model analysis in the gastrointestinal tract. World J Gastroenterol 2006; 12(18): 2870-2875

[5] Xiaona Wang and M. Q. -. Meng, Study of Frictional Properties of the Small Intestine for Design of Active Capsule Endoscope, The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, 2006, BioRob 2006, 2006, pp. 124-129, doi: 10.1109/BIOROB.2006.1639071

[6] Kim JS, Sung IH, Kim YT, Kim DE, Jang YH. Analytical Model Development for the Prediction of the Frictional Resistance of a Capsule Endoscope inside an Intestine. Proc. Inst. Mech. Eng. H. (2007) 221 (8), 837-845

[7] Kim JS, Sung IH, Kim YT, Kwon EY, Kim DE, Jang YH. Experimental investigation of frictional and viscoelastic properties of intestine for microendoscope application. Tribology Letters, Vol. 22, No. 2, May 2006, DOI: 10.1007/s11249-006-9073-0

[8] Zhou H, Alici G, Than TD, Li W. An investigation into biomechanical properties of a real intestine for design of a spiral-type robotic capsule. 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Wollongong, Australia, July 9-12, 2013

[9] Chen B, Zhou Y, Mu X. Biomechanical behaviour study of dog's small intestines. ISSN 1009-3095 Journal of Zhejiang University SCIENCE V. 3, No. 5, pp. 549-552, Nov.-Dec., 2002

[10] Accoto D, Stefanini C, Phee L, Arena A, Pernorio G, Menciassi A, Carrozza MC, Dario P. Measurements of the frictional properties of the gastrointestinal tract. Scuola Superiore Sant'Anna - MiTech Lab, via Carducci, 40 - 56127 Pisa, Italy

[11] Tan R, Liu H, Su G, Zhang Ch, Li H, Wang Y. Experimental Investigation of the Small Intestine's Viscoelasticity for the Motion of Capsule Robot. Proceedings of the 2011 IEEE, International Conference on Mechatronics and Automation, August 7-10, Beijing, China

[12] Kwon J, Park S, Kim B, Park JO. Bio-Material Property Measurement System for Locomotive Mechanism in Gastro-Intestinal Tract. Proceedings of the 2005 IEEE, International Conference on Robotics and Automation, Barcelona, Spain, April 2005

[13] Zhang Ch, Liu H. Analytical Friction Model of the Capsule Robot in the Small Intestine. Tribol Lett (2016) 64:39, DOI 10.1007/s11249-016-0774-8

[14] Higa M, Luo Y, Okuyama T, Takagi T, Shiraishi Y, Yambe T. Passive mechanical properties of large intestine under in vivo and in vito compresion. Medical Engineering & Physics 29 (2007) 840-844

Mathematical Description of the Universal IDM - some Comments and Application

Tamás Péter[1], András Háry[2], Ferenc Szauter[3], Krisztián Szabó[4], Tibor Vadvári[5], István Lakatos[3]

[1] Department of Control for Transportation and Vehicle Systems, Budapest University of Technology and Economics; Stoczek u. 2, H-1111 Budapest, Hungary; peter.tamas@mail.bme.hu

[2] ZalaZONE; Industrial Park Ltd., Dr. Michelberger Pál u. 3, H-8900 Zalaegerszeg, Hungary; andras.hary@apnb.hu

[3] Széchenyi István University, Audi Hungaria Faculty of Automotive Engineering, Department of Road and Rail Vehicles; Egyetem tér 1, H-9026 Győr, Hungary; szauter@sze.hu; lakatos@sze.hu

[4] Institute for Computer Science and Control (SZTAKI), Eötvös Loránd Research Network (ELKH); Kende u. 13-17, H-1111 Budapest, Hungary; szabo.krisztian@sztaki.hu

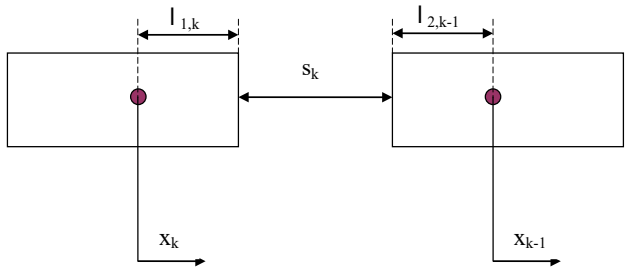
[5] University of Pannonia; Gasparich Márk u. 18/A, H-8900 Zalaegerszeg, Hungary; vadvari.tibor@zek.uni-pannon.hu

Abstract: The aim of the study is to define and mathematically describe the universal IDM. An important result of this research is that the model uses a single system of differential equations. It is able to simultaneously describe the dynamic operations of the IDM systems for all different vehicle sequences. The aim of the study is to support the driving of autonomous vehicles by taking into account the dynamic variations in the state characteristics of traffic processes. The approach used is motivated by important issues in current modelling techniques that address significant economic problems in the application of large-scale ITS network models. This also points to a new opportunity in the key area of vehicle traffic management, in the related targeted fundamental research, particularly in the analysis of traffic processes in large-scale dynamic networks.

Keywords: Universal IDM; differential equations system; dynamics-based approach; autonomous vehicles

1 Introduction

The Intelligent Driver Model (IDM) belongs to the family of adaptive cruise control (ACC) system models. It was developed in 2000 by Treiber, Hennecke and Helbing at the Transport Laboratory of the Technical University of Dresden and used by the car manufacturer BMW. For the multi-model open source road traffic simulator, Treiber and Helbing [10] use the IDM to simulate the longitudinal motion of the vehicle and this simulator also presents a lane change strategy with a software solution. The inhomogeneity of model-based single-lane traffic was investigated by Treiber et al. [13]. Treiber et al. [14], examines the stability of vehicle traffic and the parameter sensitivity of the IDM. Kesting et al. [16] propose the extension of the driver parameters of the IDM model. The authors investigate the impact of IDM-equipped vehicles on traffic flow and travel time. Jerath [4] also uses the IDM and studies the effect of adaptive cruise control on traffic processes. The results of the above works show that increasing the proportion of ACC-equipped vehicles leads to an increase in traffic efficiency by reducing travel time. Treiber and Kesting [5, 15] investigated the instability of congested traffic using IDM. The classic IDM is a chain model-like microscopic model consisting of n vehicles that describes the longitudinal dynamics of the vehicles, Figure 1 and Figure 2. Each driver looks only forward and tries to maintain an appropriate following distance. There is no overtaking in the model, i.e. the vehicles maintain their order. The first vehicle in the group and slow moving vehicles also play a key role in the model. The longitudinal dynamics of the vehicle traffic system are determined by the parameters and relationship functions of the system, and this allows the vehicles to adapt their speed to the environment (1).



$$s_k = (x_{k-1} - l_{2,k-1}) - (x_k + l_{1,k}) = (x_{k-1} - x_k) - (l_{2,k-1} + l_{1,k}) = (x_{k-1} - x_k) - l_k$$

Figure 1
The distance s_k between successive vehicles

Regarding the longitudinal dynamics, we followed the classic dynamic principle for all vehicles when creating the model. The coordinates x_k (k=1,2, …, n) describe the longitudinal movement of the vehicle's centers of gravity. In the model defined by us, the length of each k-th vehicle can be obtained from the sum of $l_{1,k}$ and $l_{2,k}$, ($L_k = l_{1,k} + l_{2,k}$). If we use the simplifying condition - that the center of gravity is at the geometric center of the length of each vehicle, then in this case $0.5\ L_k = l_{1,k}$ (k=1,2, …, n) and $s_k = (x_{k-1} - x_k) - 0.5(L_{k-1} + L_k)$.

$$\dot{v_k} = a_k \left[1 - \left(\frac{v_k}{v_k^0} \right)^{\delta} - \left(\frac{s^*(v_k, \Delta v_k}{s_k} \right)^2 \right] \tag{1}$$

For each vehicle, the motions according to their position are determined by the following system of differential equations (2) for vehicle groups 1, 2, ..., n. Derbel, O.; Peter, T.; Zebiri, H.; Mourllion, B.; Basset, M. [1, 2]. The following Derbel, O.; Peter, T.; Zebiri, H.; Mourllion, B.; For the Basset, M. [1, 2] model, we have already taken into account the more precise correlation using s*.

$s^*\left(\dot{x}_k, (\dot{x}_{k-1} - \dot{x}_k)\right) = s_0 + \dot{x}_k T + \frac{\dot{x}_k(\dot{x}_{k-1} - \dot{x}_k)}{2\sqrt{ab}}$; *($k$=1,2,…,$n$).*

This can be found in the material of Treiber, M and Kesting, A [17], which discusses the various research areas of transport processes in great detail.

$$\frac{\ddot{x_1}}{a_1} + \left(\frac{\dot{x}_1}{v_1^0} \right)^{\delta} + \left(\frac{s_0 + \dot{x}_1 T + \frac{\dot{x}_1(\dot{x}_0 - \dot{x}_1)}{2\sqrt{ab}}}{(x_0 - x_1) - l_1} \right)^2 = 1$$

$$\frac{\ddot{x_2}}{a_2} + \left(\frac{\dot{x}_2}{v_2^0} \right)^{\delta} + \left(\frac{s_0 + \dot{x}_2 T + \frac{\dot{x}_2(\dot{x}_1 - \dot{x}_2)}{2\sqrt{ab}}}{(x_1 - x_2) - l_2} \right)^2 = 1$$

$$\frac{\ddot{x_n}}{a_n} + \left(\frac{\dot{x}_n}{v_n^0} \right)^{\delta} + \left(\frac{s_0 + \dot{x}_n T + \frac{\dot{x}_n(\dot{x}_{n-1} - \dot{x}_n)}{2\sqrt{ab}}}{(x_{n-1} - x_n) - l_n} \right)^2 = 1 \tag{2}$$

Where meaning of the parameters:

a_k is the maximum acceleration of the k-th vehicle,

x_k is the position of the k-th vehicle,

$\dot{x}_k = v_k$ is the speed of the k-th vehicle,

$\ddot{x}_k$ is the acceleration of the *k*-th vehicle,

v_0^k is the desired speed of the *k*-th vehicle,

$x_{k-1} - x_k$ is the distance between the centre of gravity of the *(k-1)*-th and *k*-th vehicles,

s_k is the distance between the (k-1)-th and k-th vehicles (the vehicle length, in the case of our article, varies by vehicle, Figure 1.

$\Delta v_k = v_{k-1} - v_k$ is the difference between the speed of *(k-1)*-th and *k*-th

T Safe time headway

a the Maximum acceleration

b Comfortable deceleration

δ Acceleration exponent

s_0 Minimum distance

Modeling and research work covers a complex field and includes both microscopic and macroscopic modeling approaches, Treiber et al. [11, 12], as well as, e.g., the Generalized Velocity–Density Model based on microscopic traffic simulation, Derbel, O., Peter, T., Mourllion B., and Basset M. [3], Regarding the complex macroscopic traffic environment, the generation of the large-scale network model is the important task, for example, Péter T. and Bokor J. [7, 8, 9].

In these studies [1, 2, 3, 6], we already applied one (3) generalized, structural method to the IDM. At the same time, this is only an IDM in which the vehicle positions and vehicle numbers are the same (in the first position is vehicle 1, in the second position is vehicle 2 and in the nth position is the nth vehicle).

In our article, the procedure presented below is, how to write down and store all possible sequential IDMs in a single matrix-difference system?

$$\left\langle \underline{\underline{A}} \right\rangle^{-1} \ddot{x}(t) + \left\langle \underline{\underline{V}} \right\rangle^{-1} \underline{f_1}(\dot{x}(t)) + \left\langle \underline{\underline{S}} \right\rangle \underline{f_2}(x(t)) = \underline{1} \tag{3}$$

Where:

$$\left\langle \underline{\underline{A}} \right\rangle^{-1} = \left\langle \frac{1}{a_1}, \frac{1}{a_2}, ..., \frac{1}{a_n} \right\rangle ; \left\langle \underline{\underline{V}} \right\rangle^{-1} = \left\langle \frac{1}{v_1^4}, \frac{1}{v_2^4}, ..., \frac{1}{v_n^4} \right\rangle ; \left\langle \underline{\underline{S}} \right\rangle = \left\langle s_1^2, s_2^2, ..., s_n^2 \right\rangle$$

(In the formulas, the value of the acceleration exponent δ was set to 4.)

$$s_i = s_i(\dot{x}_{i-1}, \dot{x}_i) \qquad (i=1,2, \ldots, n).$$

$$\underline{f_1}(\dot{x}(t)) = \begin{bmatrix} \dot{x}_1^4 \\ \dot{x}_2^4 \\ \cdots \\ \dot{x}_n^4 \end{bmatrix} \qquad \underline{f_2}(x(t)) = \begin{bmatrix} \frac{1}{(x_0 - x_1 - l_1)^2} \\ \frac{1}{(x_1 - x_2 - l_2)^2} \\ \cdots \\ \frac{1}{(x_{n-1} - x_n - l_n)^2} \end{bmatrix} \qquad \underline{1} = \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix}$$

2 The Universal IDM

The model presented in this article, a mathematically, it is a complex model containing n different vehicles in every possible order. In the model, the parameter structure can be fixed, but the parameters can also be stochastic if this is necessary in connection with further investigations.

This allows for a very efficient, automatic model generation in connection with a large number of differential equation systems belonging to different IDMs.

The Universal IDM will be, a generalization of classic IDM, which still requires certain further research.

In this case, this generalization means that in a single complex mathematical model, vehicle elements i = 1,2, ..., n can freely implement overtaking strategies, on a network sector.

In our case, since we used a mathematical construction, it is advisable to continue further mathematical investigations based on this.

3 Relationship between the IDM Model and the Network Domain

For a given vehicle, the speed and the tracking distance are determined by the driver. Your decision depends on both your own perceptions and the signals sent by your vehicle from the physical environment. All of this has a crucial impact on network traffic. Accordingly, the quality of the road and the physical effects of meteorological and visual conditions determine the selectable speed range for a given vehicle density. The IDM can be used to describe the effect of dynamic

relationships between successive vehicles in a given section. At the same time, the dynamics of the motion of the IDM group is not only self-regulating, but is also determined by the control speeds of the large-scale network and network sectors [18, 19, 20, 21, 22, 23, 24, 28].

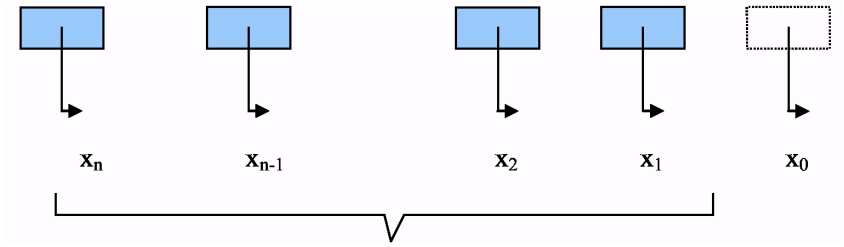


Figure 2
n elements, n! possible orders, the control function $x_0(t)$

The vehicles slow down when congestion occurs, stop when the traffic light turns red, but accelerate to the maximum speed limit after a reaction time delay if the road section ahead is clear. This is indicated in Fig. 2 by the control function $x_0(t)$, which is determined by the large-scale macroscopic network processes of each trajectory.

Based on the above, the IDM groups of all sectors are also continuously being reorganized. On this basis, it is important to emphasize that in each sector, according to reality:

1) The number of vehicles changes dynamically.

2) The order of the vehicles changes dynamically.

3) The composition of vehicles changes dynamically.

4) Overtaking manoeuvres are carried out in accordance with the parameters of the vehicles in the system taking into account the actual conditions.

5) These processes can be controlled according to the needs of the different vehicles and the traffic situation.

The usage of the model:

1) Each sector independently controls the addition and deletion of continuous input and output flows to and from the IDM operating on it.

2) Each sector independently controls and manages the input and output flows of the IDM system operating on it, and also takes into account the regulations and prohibitions resulting from the management of network traffic.

4 Mathematical Approach and Modelling Procedure

In the model, a fixed initial state is described by the vector **p**, where, for example, for n = 3, the default state is the following: p[1] = 1, p[2] = 2, p[3] = 3.

As we move forward, of course, positions may change and the value of the vector coordinate p_i will be, for example, k.

The meaning of p_i = k is that the *k*-th vehicle is in the *i*-th position in the queue at the given time (i = 1, 2, …, n; k = 1,2, …, n;).

In the model, the existence of a direct relationship between two vehicles is determined by the values of the elements of matrix **U** defined by us. The existence of a relationship is represented by the value 1, while the fact that there is no relationship between them is defined by the value 0.

Consequently, these relationships are contained in the matrix u[i, k], (i =1,2, …, n; k =1,2, …, n), in which in each *i*-th row, only u[i, p[i]] = 1, thus u[i, k] = 1 and the other elements of the *i*-th row are zero (i = 1,2, …, n).

In the system of differential equations of the universal IDM, the elements of the matrix u[i, k] are used by the matrices **A** and **B**. Matrix **A** takes into account the maximum acceleration a_i of the *i*-th vehicle, while matrix **B** takes into account the desired speed v_i of the *i*-th vehicle.

In the case of the system of differential equations the 3-dimensional matrix k[j, p, i] is used in the vector ***f***. The vector ***f*** determines the relations for the quotients of the squared distances between the two vehicles or determines the non-existence of these relations.

Let us consider first the first equation of the system of differential equations. Then the value of the matrix k[j, p, i] will be 1 only in the case (4) when:

$$k[0, p[1],1]=1 \tag{4}$$

Namely, it is the value of p[1] that determines which *k*-th vehicle is at the front of the queue and follows the control signal $x_0(t)$.

Accordingly, the value of the other elements in the sum is zero.

Next, let us examine the values of the elements of the matrix k[j, p, i] in the second equation of the system of differential equations.

Here, a value of 1 only applies to vehicles with serial numbers at positions p_1 and p_2, according to the following conditions:

$$\begin{aligned} &\text{if } ((p[1] < p[2])) \text{ then } k[p[1], p[2], 2] =1 \\ &\text{if } ((p[1] > p[2])) \text{ then } k[p[2], p[1], 2] =1 \end{aligned} \tag{5}$$

the other elements in the sum are zero, and so on.

In the last line of the system of differential equations, 1 value also occurs in only two cases. Exactly, only for vehicles with serial numbers at positions p_{n-1} and p_n, under the following conditions:

if ((p[n-1]<p[n])) then k[p[n-1],p[n],n] =1; (6)

if ((p[n-1]>p[n])) then k[p[n],p[n-1],n] =1

the other elements in the sum are zero.

One result of this method is that the IDM for all permutations of the n-element vehicle group can be summarized in a single model. Based on this, the IDM for any vehicle sequence can be written by appropriately setting the elements of the matrices u[i, k] and k[j, p, i]. This model definition thus allows for complex modeling of definitively formed sequences based on a single model.

Another important result of the method is that it integrated all possible n-element IDMs into a single complex model.

In this way, it also prepares the possibility of a mathematical analysis of structural changes occurring during system transitions.

5 The System of Differential Equations of the Universal IDM for n Vehicles

The requirements are described by the following system of matrix differential equations:

$$A \cdot \ddot{x} + B \cdot \dot{x}^4 + f(x) = e(t) \tag{7}$$

(In the formulas, the value of the acceleration exponent δ was set to 4. Naturally, 4 can be replaced with δ and the acceleration exponent can be used in general.)

Where, on the left side of the above system of matrix differential equations, the products, are column vectors. The elements of the matrices **A** and **B** comprise the matrix elements $u_{i,j}$ discussed above, as follows:

$$A= [u_{i,j} 1/a_j]; B= [u_{i,j}\ 1/v_j^4]$$

In mathematics, this is the **Hadamard product**, also known as the **Schur product**, which is the product of the elements in the same place in two matrices, resulting in a matrix of the same dimension as the two matrices. It is named after the French mathematician Jacques Hadamard and the German mathematician Issai Schur. The vector ***f*** contains the matrix elements $k_{j,\,p,\,i}$.

Where:

$$A=\begin{bmatrix} \frac{u_{11}}{a_1} & \frac{u_{12}}{a_2} & \dots & \frac{u_{1n}}{a_n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \frac{u_{ij}}{a_j} & \dots \\ \frac{u_{n1}}{a_1} & \frac{u_{n2}}{a_2} & \dots & \frac{u_{nn}}{a_n} \end{bmatrix}; \quad B=\begin{bmatrix} \frac{u_{11}}{v_1^4} & \frac{u_{12}}{v_2^4} & \dots & \frac{u_{1n}}{v_n^4} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \frac{u_{ij}}{v_j^4} & \dots \\ \frac{u_{n1}}{v_1^4} & \frac{u_{n2}}{v_2^4} & \dots & \frac{u_{nn}}{v_n^4} \end{bmatrix};$$

$$f=\begin{bmatrix} \sum_{p=1}^{n} \frac{k_{0,p,1}\cdot s_{0,p}{}^2}{(x_0-x_p-l_p)^2} \\ \dots \\ \sum_{j=1}^{n-1}\sum_{p=j+1}^{n} \frac{k_{j,p,i}\cdot s_{j,p}{}^2}{(x_j-x_p-l_p)^2} \\ \sum_{j=1}^{n-1}\sum_{p=j+1}^{n} \frac{k_{j,p,n}\cdot s_{j,p}{}^2}{(x_j-x_p-l_p)^2} \end{bmatrix}$$

$x\in\Re^n$ is the state characteristic vector of the vehicle positions, where n is the number of vehicles,

$\dot{x}\in\Re^n$, $\ddot{x}\in\Re^n$ are the vectors of vehicle speeds and accelerations

$A, B\in\Re^{nxn}$ $f,e\in\Re^n$

In a more concise notation, the matrices and vectors in the system of equations can be summarized as follows:

$$A=\begin{bmatrix} \frac{u_{ij}}{a_j} \end{bmatrix}; \quad B=\begin{bmatrix} \frac{u_{ij}}{v_j^4} \end{bmatrix} \quad f=\begin{bmatrix} f_i \end{bmatrix}$$

$(i=1, 2, \dots, n; j=1, 2, \dots, n)$

for i=1:

$$\sum_{p=1}^{n} \frac{k_{0,p,1}\cdot s_{0,p}{}^2}{(x_0-x_p-1_p)^2}$$

Where:

$$s_{0,p}^2=\left(s_0+\dot{x}_pT+\frac{\dot{x}_p(\dot{x}_0-\dot{x}_p)}{2\sqrt{ab}}\right)^2;\ (p=1,2,\dots,n);$$

for i>1:

$$f_i = \sum_{j=1}^{n-1} \sum_{p=j+1}^{n} \frac{k_{j,p,i} \cdot {s_{j,p}}^2}{(x_j - x_p - 1_p)^2}$$

Where:

$$s_{j,p}^2 = \left(s_0 + \dot{x}_p T + \frac{\dot{x}_p (\dot{x}_j - \dot{x}_p)}{2\sqrt{ab}} \right)^2 ; (j=1,2,\ldots,n\text{-}1; p=j+1,\ldots,n);$$

The total number of all IDM is n! Let us consider, as an example, the collective system of differential equations containing all sequences for n = 3:

$$\begin{bmatrix} \frac{u_{1,1}}{a_1} & \frac{u_{1,2}}{a_2} & \frac{u_{1,3}}{a_3} \\ \frac{u_{2,1}}{a_1} & \frac{u_{2,2}}{a_2} & \frac{u_{2,3}}{a_3} \\ \frac{u_{3,1}}{a_1} & \frac{u_{3,2}}{a_2} & \frac{u_{3,3}}{a_3} \end{bmatrix} \begin{bmatrix} \frac{\mathrm{d}^2}{\mathrm{d}t^2} x1(t) \\ \frac{\mathrm{d}^2}{\mathrm{d}t^2} x2(t) \\ \frac{\mathrm{d}^2}{\mathrm{d}t^2} x3(t) \end{bmatrix} + \begin{bmatrix} \frac{u_{1,1}}{v_1^4} & \frac{u_{1,2}}{v_2^4} & \frac{u_{1,3}}{v_3^4} \\ \frac{u_{2,1}}{v_1^4} & \frac{u_{2,2}}{v_2^4} & \frac{u_{2,3}}{v_3^4} \\ \frac{u_{3,1}}{v_1^4} & \frac{u_{3,2}}{v_2^4} & \frac{u_{3,3}}{v_3^4} \end{bmatrix} \begin{bmatrix} \left(\frac{\mathrm{d}}{\mathrm{d}t} x1(t) \right)^4 \\ \left(\frac{\mathrm{d}}{\mathrm{d}t} x2(t) \right)^4 \\ \left(\frac{\mathrm{d}}{\mathrm{d}t} x3(t) \right)^4 \end{bmatrix}$$
$$+ \begin{bmatrix} \frac{k_{0,1,1} s_{0,1}^2}{(x_0 - x_1 - l_1)^2} + \frac{k_{0,2,1} s_{0,2}^2}{(x_0 - x_2 - l_2)^2} + \frac{k_{0,3,1} s_{0,3}^2}{(x_0 - x_3 - l_3)^2} \\ \frac{k_{1,2,2} s_{1,2}^2}{(x_1 - x_2 - l_2)^2} + \frac{k_{1,3,2} s_{1,3}^2}{(x_1 - x_3 - l_3)^2} + \frac{k_{2,3,2} s_{2,3}^2}{(x_2 - x_3 - l_3)^2} \\ \frac{k_{1,2,3} s_{1,2}^2}{(x_1 - x_2 - l_2)^2} + \frac{k_{1,3,3} s_{1,3}^2}{(x_1 - x_3 - l_3)^2} + \frac{k_{2,3,3} s_{2,3}^2}{(x_2 - x_3 - l_3)^2} \end{bmatrix} = \begin{bmatrix} e(t) \\ e(t) \\ e(t) \end{bmatrix} \quad (8)$$

In total, six sequences can be generated for each initial case. All the cases that can be automatically generated by computer algebra are described below.

Case No.1: where the order of the vehicles is 1, 2, 3, contained in the vector p:

$$p^1=1,\ p^2=2,\ p^3=3$$

In this case, the calculated values of $u_{i,j}$ are as follows for u[i,p[i]]:=1; (i = 1,2,3):

$$u^{1,1}=1,\ u^{2,2}=1,\ u^{3,3}=1$$

The following algorithm is used to calculate the values of **k**[j,p,i], for **k**[0,p[1],1]:=1:

```
if ((p[1] <p[2])) then k[p[1],p[2],2]:=1; end; if ((p[1]>p[2])) then
                    k[p[2],p[1],2]:=1; end if;
if ((p[2] <p[3])) then k[p[2],p[3],3]:=1; end; if ((p[2]>p[3])) then
                    k[p[3],p[2],3]:=1; end if;
```
(9)

The values of $k_{i,j,l}$ calculated using the algorithm are as follows:

$$k^{0,1,1}=1,\ k^{1,2,2}=1,\ k^{2,3,3}=1$$

Based on the above, the following final system of differential equations was determined in case No. 1 (10):

$$\begin{bmatrix} \frac{\frac{\mathrm{d}^2}{\mathrm{d}t^2}x1(t)}{a_1} \\ \frac{\frac{\mathrm{d}^2}{\mathrm{d}t^2}x2(t)}{a_2} \\ \frac{\frac{\mathrm{d}^2}{\mathrm{d}t^2}x3(t)}{a_3} \end{bmatrix} + \begin{bmatrix} \frac{\left(\frac{\mathrm{d}}{\mathrm{d}t}x1(t)\right)^4}{v_1^4} \\ \frac{\left(\frac{\mathrm{d}}{\mathrm{d}t}x2(t)\right)^4}{v_2^4} \\ \frac{\left(\frac{\mathrm{d}}{\mathrm{d}t}x3(t)\right)^4}{v_3^4} \end{bmatrix} + \begin{bmatrix} \frac{s_{0,1}^2}{(x_0-x_1-l_1)^2} \\ \frac{s_{1,2}^2}{(x_1-x_2-l_2)^2} \\ \frac{s_{2,3}^2}{(x_2-x_3-l_3)^2} \end{bmatrix} = \boldsymbol{e}(t) \tag{10}$$

Following the above algorithm, the computer algebraic method automatically provides the further results.

Case No. 2: the order of the vehicles is 1, 3, 2. The calculated values of $u_{i,j}$ and $k_{i,j,l}$ are:

$$p^1=1,\ p^2=3,\ p^3=2$$

$$u^{1,1}=1,\ u^{2,3}=1,\ u^{3,2}=1$$

$$k^{0,1,1}=1,\ k^{1,3,2}=1,\ k^{2,3,3}=1$$

In case No. 2, the final system of differential equations is as follows:

$$\begin{bmatrix} \frac{\frac{\mathrm{d}^2}{\mathrm{d}t^2}x1(t)}{a_1} \\ \frac{\frac{\mathrm{d}^2}{\mathrm{d}t^2}x3(t)}{a_3} \\ \frac{\frac{\mathrm{d}^2}{\mathrm{d}t^2}x2(t)}{a_2} \end{bmatrix} + \begin{bmatrix} \frac{\left(\frac{\mathrm{d}}{\mathrm{d}t}x1(t)\right)^4}{v_1^4} \\ \frac{\left(\frac{\mathrm{d}}{\mathrm{d}t}x3(t)\right)^4}{v_3^4} \\ \frac{\left(\frac{\mathrm{d}}{\mathrm{d}t}x2(t)\right)^4}{v_2^4} \end{bmatrix} + \begin{bmatrix} \frac{s_{0,1}^2}{(x_0-x_1-l_1)^2} \\ \frac{s_{1,3}^2}{(x_1-x_3-l_3)^2} \\ \frac{s_{2,3}^2}{(x_2-x_3-l_3)^2} \end{bmatrix} = \boldsymbol{e}(t) \tag{11}$$

Case No. 3: the order of the vehicles is 2, 1, 3. The calculated values of $u_{i,j}$ and $k_{i,j,l}$ are:

$$p^1=2\, p^2=1, p^3=3$$

$$u^{1,2}=1,\ u^{2,1}=1,\ u^{3,3}=1$$

$$k^{0,2,1}=1,\ k^{1,2,2}=1,\ k^{1,3,3}=1$$

In case No. 3, the final system of differential equations is as follows:

$$\begin{bmatrix} \frac{\frac{d^2}{dt^2}x2(t)}{a_2} \\ \frac{\frac{d^2}{dt^2}x1(t)}{a_1} \\ \frac{\frac{d^2}{dt^2}x3(t)}{a_3} \end{bmatrix} + \begin{bmatrix} \frac{\left(\frac{d}{dt}x2(t)\right)^4}{v_2^4} \\ \frac{\left(\frac{d}{dt}x1(t)\right)^4}{v_1^4} \\ \frac{\left(\frac{d}{dt}x3(t)\right)^4}{v_3^4} \end{bmatrix} + \begin{bmatrix} \frac{s_{0,2}^2}{(x_0-x_2-l_2)^2} \\ \frac{s_{1,2}^2}{(x_1-x_2-l_2)^2} \\ \frac{s_{1,3}^2}{(x_1-x_3-l_3)^2} \end{bmatrix} = \boldsymbol{e}(t) \qquad (12)$$

Case No. 4: the order of the vehicles is 2, 3, 1. The calculated values of $u_{i,j}$ and $k_{i,j,l}$ are:

$$p^1=2,\ p^2=3, p^3=1$$

$$u^{1,2}=1,\ u^{2,3}=1,\ u^{3,1}=1$$

$$k^{0,2,1}=1,\ k^{2,3,2}=1,\ k^{1,3,3}=1$$

In case No. 4, the final system of differential equations is as follows:

$$\begin{bmatrix} \frac{\frac{d^2}{dt^2}x2(t)}{a_2} \\ \frac{\frac{d^2}{dt^2}x3(t)}{a_3} \\ \frac{\frac{d^2}{dt^2}x1(t)}{a_1} \end{bmatrix} + \begin{bmatrix} \frac{\left(\frac{d}{dt}x2(t)\right)^4}{v_2^4} \\ \frac{\left(\frac{d}{dt}x3(t)\right)^4}{v_3^4} \\ \frac{\left(\frac{d}{dt}x1(t)\right)^4}{v_1^4} \end{bmatrix} + \begin{bmatrix} \frac{s_{0,2}^2}{(x_0-x_2-l_2)^2} \\ \frac{s_{2,3}^2}{(x_2-x_3-l_3)^2} \\ \frac{s_{1,3}^2}{(x_1-x_3-l_3)^2} \end{bmatrix} = \boldsymbol{e}(t) \qquad (13)$$

Case No. 5: the order of the vehicles is 3, 1, 2. The calculated values of $u_{i,j}$ and $k_{i,j,l}$ are:

$$p^1=3,\ p^2=1, p^3=2$$

$$u^{1,3}=1,\ u^{2,1}=1,\ u^{3,2}=1$$

$$k^{0,3,1}=1,\ k^{1,3,2}=1,\ k^{1,2,3}=1$$

In case No. 5, the final system of differential equations is as follows:

$$\begin{bmatrix} \frac{\frac{d^2}{dt^2}x3(t)}{a_3} \\ \frac{\frac{d^2}{dt^2}x1(t)}{a_1} \\ \frac{\frac{d^2}{dt^2}x2(t)}{a_2} \end{bmatrix} + \begin{bmatrix} \frac{\left(\frac{d}{dt}x3(t)\right)^4}{v_3^4} \\ \frac{\left(\frac{d}{dt}x1(t)\right)^4}{v_1^4} \\ \frac{\left(\frac{d}{dt}x2(t)\right)^4}{v_2^4} \end{bmatrix} + \begin{bmatrix} \frac{s_{0,3}^2}{(x_0-x_3-l_3)^2} \\ \frac{s_{1,3}^2}{(x_1-x_3-l_3)^2} \\ \frac{s_{1,2}^2}{(x_1-x_2-l_2)^2} \end{bmatrix} = \boldsymbol{e}(t) \tag{14}$$

Case No. 6: the order of the vehicles is 3, 2, 1. The calculated values of $u_{i,j}$ and $k_{i,j,l}$ are:

$$p^1=3,\ p^2=2,\ p^3=1$$

$$u^{1,3}=1,\ u^{2,2}=1,\ u^{3,1}=1$$

$$k^{0,3,1}=1,\ k^{2,3,2}=1,\ k^{1,2,3}=1$$

In case No. 6, the final system of differential equations is as follows:

$$\begin{bmatrix} \frac{\frac{d^2}{dt^2}x3(t)}{a_3} \\ \frac{\frac{d^2}{dt^2}x2(t)}{a_2} \\ \frac{\frac{d^2}{dt^2}x1(t)}{a_1} \end{bmatrix} + \begin{bmatrix} \frac{\left(\frac{d}{dt}x3(t)\right)^4}{v_3^4} \\ \frac{\left(\frac{d}{dt}x2(t)\right)^4}{v_2^4} \\ \frac{\left(\frac{d}{dt}x1(t)\right)^4}{v_1^4} \end{bmatrix} + \begin{bmatrix} \frac{s_{0,3}^2}{(x_0-x_3-l_3)^2} \\ \frac{s_{2,3}^2}{(x_2-x_3-l_3)^2} \\ \frac{s_{1,2}^2}{(x_1-x_2-l_2)^2} \end{bmatrix} = \boldsymbol{e}(t) \tag{15}$$

The above system of equations therefore describes an already formed and constant series over a certain period of time.

Conclusions

It can be concluded that the application of IDM chain models in itself has provided and is currently providing opportunities for many useful investigations related to traffic processes. In this regard, the automatic generation of a large number of these models provides a very useful additional contribution to these studies, model studies have been carried out, e.g., [30], [34]. The specified model,

mathematically, it is a complex model containing n different vehicles in every possible order. In the model, the parameter structure can be fixed, but the parameters can also be stochastic, if this is necessary in connection with further investigations. The (7) is suitable for generating IDMs for different n-element vehicle lines since the entire connection system is located and stored in a single complex system of differential equations. This allows for a very efficient, automatic model generation in connection with a large number of differential equation systems belonging to different IDMs.

Based on the formulation of the principles, it is important to determine which objectives should be applied under the given traffic conditions. These can be: optimal proceeding through intersections, optimal energy consumption in traffic processes and the related optimal CO2 emissions, optimal environmental impact, and rapid transfer of vehicle convoys through the network domain [25, 26, 27, 29, 31, 32, 33]. In a more general approach, the parameters of the IDM system are transport system parameters. The vehicle dynamics parameters of the IDM can be investigated by an appropriately chosen rewriting of the mathematical model. The chosen model structure is also suitable for the analysis of the multi-mass vibrational dynamics model, using purely vibration theory concepts [6].

In the field of algorithms and programming related to overtaking and lane changes, there are indeed many excellent procedures. In our case, since we used a mathematical construction, it is advisable to continue further mathematical investigations based on this. The change in structure during system transitions is a very interesting problem and it really requires further important research, e.g. [35].

Acknowledgement

The research presented in this paper was carried out as part of the TKP2021-NKTA-48 a Ministry of Technology and Industry National Research, With support from the Development and Innovation Fund, a Funded by the TKP2021-NKTA tender program" was realized.

References

[1] Derbel, O.; Peter, T.; Zebiri, H.; Mourllion, B.; Basset, M. (2012) Modified intelligent driver model, *Periodica Polytechnica Transportation Engineering* 40(2): 53-60, https://doi.org/10.3311/pp.tr.2012-2.02

[2] Derbel, O.; Peter, T.; Zebiri, H.; Mourllion, B.; Basset, M. (2013) Modified intelligent driver model for driver safety and traffic stability improvement, *IFAC Proceedings Volumes* 46(21): 744-749, https://doi.org/10.3182/20130904-4-JP-2042.00132

[3] Derbel, O., Péter, T., Mourllion B., & Basset M. (2017) Generalized Velocity–Density Model based on microscopic traffic simulation, Transport, DOI: 10.3846/16484142.2017.1292950 To link to this article: http://dx.doi.org/10.3846/16484142.2017.1292950 ISSN: 1648-4142 (Print) 1648-3480; http://www.tandfonline.com/loi/tran20

[4] Jerath, K., (2010) Impact of adaptive cruise control on the formation of self-organized traffic jams on highway. Master's thesis, The Pennsylvania State University.The Graduate School. Department of Mechanical and Nuclear Engineering

[5] Kesting, A., Treiber, M., Helbing, D., (2008) Agents for traffic simulation. Physics and Society 11, 325-356

[6] T., Péter; I., Lakatos (2019) Vehicle dynamic-based approach for the optimization of traffic parameters of the Intelligent Driver Model (IDM) and for the support of autonomous vehicles' driving ability Acta Polytechnica Hungarica 16: 3 pp. 121-142, 22 p. (2019)

[7] Péter T, and Bokor J (2010) Research for the modelling and control of traffic, In: Scientific Society for Mechanical Engineering, 33rd Fisita-World Automotive Congress: Proceedings, Budapest, Hungary, 2010.05.30-2010.06.04. Budapest: GTE, 2010. pp. 66-73 (ISBN:978-963-9058-28-6)

[8] Péter and Bokor J (2011) New road traffic networks models for control, GSTF International Journal on Computing, Vol. 1, Number 2, pp. 227-232, DOI: 10.5176_2010-2283_1.2.65 February 2011

[9] Péter, T (2012) Modeling nonlinear road traffic networks for junction control, International Journal of Applied Mathematics and Computer Science (AMCS), 2012, Vol. 22, No. 3, pp. 723-732, DOI: 10.2478/v1006-012-0054-1

[10] Treiber, M., Helbing, D., (2002) Realistische mikrosimulation von straenverkehr mit einem einfachen modell. In: Symposium "Simulationstechnik ASIM

[11] Treiber, M.; Hennecke, A.; Helbing, D. (2000a) Congested traffic states in empirical observations and microscopic simulations, *Physical Review E* 62(2): 1805-1824, https://doi.org/10.1103/PhysRevE.62.1805

[12] Treiber, M.; Hennecke, A.; Helbing, D. (2000b) Microscopic simulation of congested traffic, in D. Helbing, H. J. Herrmann, M. Schreckenberg, D. E. Wolf (Eds.). *Traffic and Granular Flow'99: Social, Traffic, and Granular Dynamics*, 365-376, https://doi.org/10.1007/978-3-642-59751-0_36

[13] Treiber, M., Hennecke, A., Helbing, D., (2004) Microscopic simulation of congested traffic. Physical Review E 62, 1805-1824

[14] Treiber, M., Hennecke, A., Helbing, D., (2006) Delays, inaccuracies and anticipation in microscopic traffic models. Physica A 360, 71-88

[15] Treiber, M., Kesting, A., (2011) Evidence of convective instability in congested traffic flow: A systematic empirical and theoretical investigation. Procedia Social and Behavioral Sciences 17, 698-716

[16] Kesting, A., Treiber, M., Helbing, D. (2008) Agents for traffic simulation. *Physics and Society* 11, 325-356

[17] Martin Treiber, Arne Kesting (2013) Traffic Flow Dynamics, Data, Models and Simulation, ISBN 978-3-642-32459-8; ISBN 978-3-642-32460-4 (eBook) DOI 10.1007/978-3-642-32460-4 Springer-Verlag Berlin Heidelberg 2013, p. 503

[18] Péter T, and Bokor J. (2010.1) Research for the modelling and control of traffic, In: Scientific Society for Mechanical Engineering ,*33rd Fisita-World Automotive Congress*: Proceedings, Budapest, Hungary, 2010.05.30-2010.06.04. Budapest: GTE, 2010, pp. 66-73 (ISBN:978-963-9058-28-6)

[19] Péter T, and Bokor J. (2010.2) Modeling road traffic networks for control. *Annual international conference on network technologies & communications: NTC 2010*. Thailand, 2010.11.30-2010.11.30. pp. 18-22, Paper 21 (ISBN:978-981-08-7654-8)

[20] Péter T. and Bokor J. (2011) New road traffic networks models for control, *GSTF International Journal on Computing*, Vol. 1, Number 2, pp. 227-232, DOI: 10.5176_2010-2283_1.2.65 February 2011

[21] Péter T, Fazekas S. (2014) Determination of vehicle density of inputs and outputs and model validation for the analysis of network traffic processes *Periodica Polytechnica Transportation Engineering*, 42:(1) *pp. 53-61,* (2014) (Budapest University of Technology and Economics)

[22] Pokorádi, L. (2018) Graph model-based analysis of technical systems *IOP Conf. Series: Materials Science and Engineering 393* (2018) 012007, pp. 1-9, doi:10.1088/1757-899X/393/1/012007

[23] Pokorádi, L., (2018) Methodology of Advanced Graph Model-based Vehicle Systems' Analysis In: Szakál, Anikó (editor) IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI 2018) Budapest, IEEE Hungary Section (2018) pp. 325-328, 4 p.

[24] Pokorádi, L., Gáti, J., (2018) Markovian Model-based Sensitivity Analysis of Maintenance System In: Anikó, Szakál (editor) IEEE 16th International Symposium on Intelligent Systems and Informatics: SISY 2018 Budapest, IEEE Hungary Section (2018) pp. 117-121, 5 p.

[25] Pokorádi, L. Molnár, B., (2010) Illustration of the Monte-Carlo simulation, Szolnok Scientific Publications XIV pp. 1-12, 12 p. (2010)

[26] Pokorádi L., (2002) Fuzzy Logic-Based Risk Assessment, Academic and Applied Research in Military Science 1588-8789 1788-0017 I 1 63-73, 1092067 (2002)

[27] Takács,_Á., Drexler, D., A., Galambos, P., Rudas,_I., J., Haidegger, T. (2018) Assessment and Standardization of Autonomous Vehicles *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)* 21-23 June., Las Palmas de Gran Canaria, Spain, pp. 185-192, ISSN: 1543-9259, DOI: 10.1109/INES.2018.8523899

[28] Robert Pethes and Levente Kovács, (2020) Voting to the Link: a Static Network Formation Model, Acta Polytechnica Hungarica Vol. 17, No. 3, 2020, pp. 207-228

[29] Tamás D. Nagy; Nikita Ukhrenkov; Daniel A. Drexler; Árpád Takács; Tamás Haidegger (2019) Enabling quantitative analysis of situation awareness: system architecture for autonomous vehicle handover studies Publisher:IEEE; SMC 2019: 904-908; [IEEE 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) - Bari, Italy (2019.10.6-2019.10.9)]

[30] Shobhit Gupta, Shreshta R. Deshpande, Punit Tulpule, Marcello Canova, Giorgio Rizzoni (2019) An Enhanced Driver Model for Evaluating Fuel Economy on Real-World Routes; Center For Automotive Research, The Ohio State University, Columbus, OH 43212 USA. 9th IFAC Symposium on Advances in Automotive Control AAC 2019 Orléans, France, 23-27 June 2019 Edited by Dominique Nelson-Gruel Volume 52, Issue 5, Pages 574-579 (2019)

[31] A. Titrik, I. Lakatos, D. Czegledi (2015): Saturation Optimization of Selective Waste Collection Vehicles Based on Real-Time Info-Communication System. ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference Adam Titrik (2016)

[32] A. Titrik. Sign-in-time Based Infocommunication System for Collecting Selective Waste. (2016) Periodica Polytechnica Transportation Engineering 0303-7800 1587-3811 44 1 1-4

[33] I. Lakatos, A. Titrik T. Orban (2011): Data determination of an internal combustion engine for model set-up. (2011) Hungarian Journal of Industry and Chemistry 0133-0276 2450-5102 39 1 35-40

[34] Yun Li , Shengrui Zhang , Yingjiu Pan , Bei Zhou , and Yanan Peng (2023) Exploring the Stability and Capacity Characteristics of Mixed Traffic Flow with Autonomous and Human-Driven Vehicles considering Aggressive Driving; Hindawi Journal of Advanced Transportation. Volume 2023, Article ID 2578690, 21 pages https://doi.org/10.1155/2023/2578690

[35] Yiling You (2022) Mathematical Models and Control Algorithms for Traffic Automation. Dissertation in Applied Mathematics, p. 101, University of California, Berkeley, 2022

Operations Research in Online Environment Using ICT in Higher Technical Education

Henrieta Hrablik Chovanová, Helena Fidlerová, Dagmar Babčanová, Daniel Banáš*

Slovak University of Technology in Bratislava, Faculty of Materials Science and Technology in Trnava, Jana Bottu 25, 917 24 Trnava, Slovakia
henrieta.chovanova@stuba.sk; helena.fidlerova@stuba.sk; dagmar.babcanova@stuba.sk; daniel.banas@stuba.sk

Abstract: Operations research is now in the renaissance because organizations are optimizing their processes using analytical methods and different research approaches to improve the managing of processes/tasks/problems. Operations research, requiring analytical, systems thinking, the ability of abstract thinking, so it is considered by its complexity and the scope of required knowledge to be a subject that is a challenge to master for students of higher technical education. The objective of the paper is to analyze and compare the situation in the teaching of operations research, identify the prerequisites for the development of analytical thinking of future graduates, and predict the possibilities for improving the education of operations research during distance learning. Part of the paper deals with the description and comparison of areas and methods of operations research taught at technical higher education institutions (HEIs) and universities in Slovakia and the Czech Republic, including the use of software support for calculating and learning the curriculum. The analysis considers results about teaching subjects related to the operations research, comparison of the form, methods, research areas, and software used for teaching at Slovak and Czech HEIs and universities, and semi-structured interviews/surveys among students about the quality of education process identified areas where there is potential for improvement. At the same time, the results of distance education were analyzed the degree of student satisfaction, awareness, comprehensiveness to conduct the new possibilities and threats of distance education. The statistical evaluation of data indicates strong relationship between the two variables: satisfaction of students and awareness of students about online education.

Keywords: operations research; quantitative methods; higher technical education; education process; awareness; online learning

1 Introduction

The precision of operations research methods brings to the decision-making certainty and substantiation of the decisions by measuring, analyzing, and interpreting the data. Graduates of technical disciplines need to master analytical systems thinking, and mastering methods of operations research (further OR) can help them. The acquired knowledge and competencies are used by engineers and managers in decision-making tasks [1].

Nevertheless, according to research findings, there are several issues in the form of teaching method, using software and content of teaching OR and related subjects, and therefore the role of the paper is to examine the situation in the field of technical higher education. Identification of possibilities in the mentioned area enables prediction of the future challenges in improving education in OR.

Several authors, including [2, 3, 4, 5, 6] claim that the use of ICT in teaching subjects including OR, significantly helps students to understand concepts, basic principles, and procedures for solving assigned tasks/defined problems.

Information and communication technology (ICT) means an important source of innovations and improvements for many sectors around the world. We can say that in the field of technical education, ICT applications are a strategic component of the success and enhancement of the education process of higher education institutions and universities [7, 8]. Therefore, government and stakeholders have invested over the last two decades in the ICT adoption in the education system, as stated by the authors [4, 5]. Most universities that have fully adopted ICT have made significant progress in the application of ICT to improve learning methods, education, research and development [4].

Education has been one of the major areas disrupted by the COVID-19 pandemic. [9] The rapid transition of courses to an emergency remote teaching and learning format at the onset of the COVID-19 pandemic in early 2020 created challenges across the university landscape for faculty and students and, inevitably, affects the future of higher education. [10, 11] Studies from the pre-Covid period and also studies conducted during the pandemic [12, 13, 14, 15, 16] approve the wide application of ICT as support tools for teaching in higher education. Another study [17] provides evidence that students prefer ICT-based learning and (they?) believe that knowledge and general learning skills acquired through use of ICT would not be achieved by traditional teaching.

Today's learners expect educators to effectively use information technologies in the classroom. [18, 19] Educators and students had to adapt in short time to the new reality and learn to work with new digital tools and technologies including *Blackboard, Moodle, Canvas, Adobe Connect, Zoom, MS Teams and Panopto*. Teaching and learning activities used in new virtual environment included online lectures and teamwork, the use of a virtual whiteboard, providing feedback to

students in real-time, also recording the lecture, adding comments to the presentation for many academics. [20, 21, 22]

2 Theoretical Background - Operation Research

As early as 1968, OR was defined by Jakff and Sasienio as follows: "Operations research can be considered the application of a scientific method by interdisciplinary teams to problems involving the control of organized (man-machine) systems to provide solutions that best serve the purpose of the organization as a whole." [23]

At present, OR is a discipline focused on solving management problems using mathematical models and methods, while ICT is used to solve "problems". In the United States, it is often referred to directly as "management science." In the article, the term OR will be understood as a scientific discipline, the subject of which is the study and analysis of operations and processes that take place or are planned in a particular organizational unit (enterprise, plant, workshop), the study and analysis of these operations are most often carried out using mathematical modelling." [24]

According to [25], Operations Research (OR) is a branch of applied mathematics and, thanks to its interdisciplinary nature, OR finds application in several areas, such as logistics, production and equipment planning, marketing and finance. We also agree with the authors that: “OR can show the possible connections between mathematics and the real world, which can support the students’ learning process, helping them to reason and develop problem-solving and analytical skills. Moreover, a positive attitude towards mathematics could foster students *to pursue a career in scientific disciplines or to continue the study at the university level*”.

OR has a broader scope, using methods and techniques from other mathematical sciences, such as statistical analysis, mathematical modelling, predictive analysis, and mathematical optimization, makes it possible to indicate optimal solutions to complex decision problems. Due to its comprehensive and practical focus on practical application, OR has overlapped with other disciplines, such as industrial engineering, logistics, and operations management.

3 Methodology

The research aimed to find out the specifics of teaching OR at an institution of higher technical education, to describe the issues and situation in this field with a focus on OR as a key subject of the master study, to evaluate the relationship between satisfaction, awareness, and fullness during online teaching.

The contribution of this paper is to identify opportunities for improvement and the challenges/potential risks of online education in the field of OR. The article aims to analyze the specifics of online teaching with a focus on OR at higher technical education institutions, including technical universities.

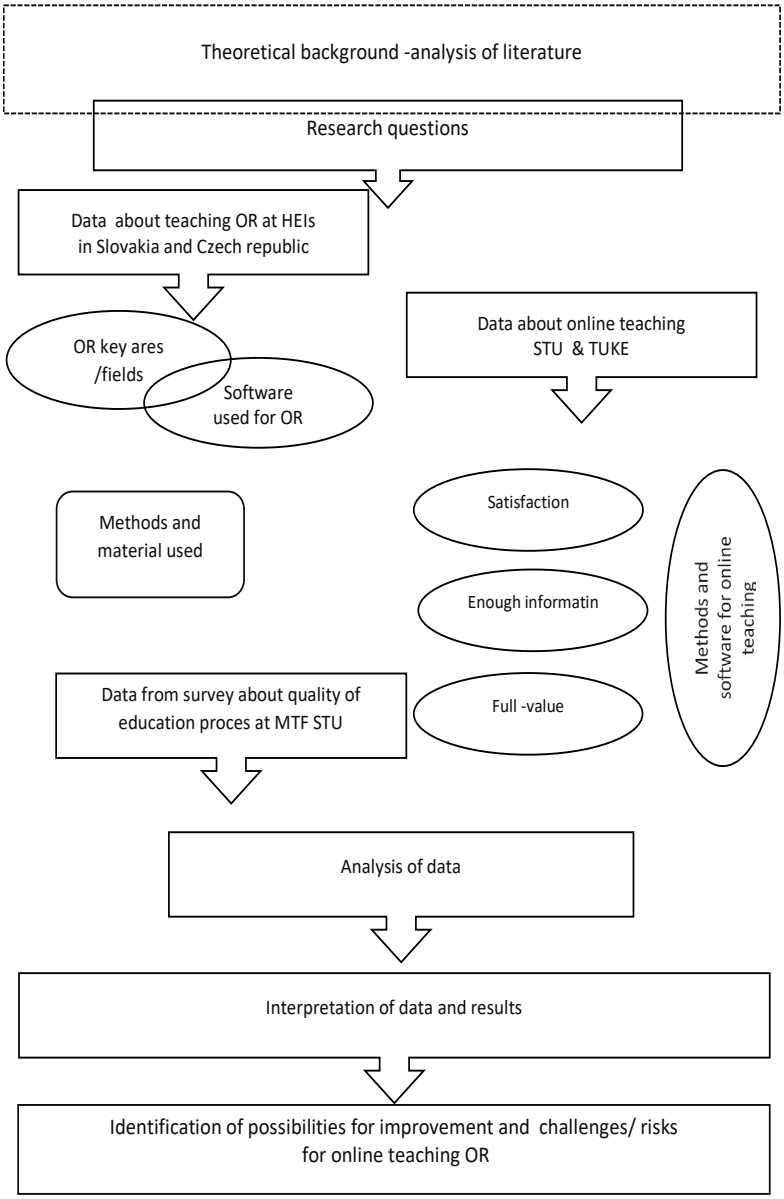


Figure 1
Analytical framework – methodology (drawn by authors)

The theoretical background has been introduced by a review of the available literature regarding OR, online teaching, and competencies. Subsequently, after identifying the gap, the research questions were claimed as follows (Fig. 1):

- Which are the key methods and key areas in OR?
- What kind of software is used to teach OR?
- What is the quality of education and the satisfaction of the students who are enrolled in subject OR?
- Can be expected the same value in online teaching as face-to-face teaching?
- What affects students' learning satisfaction in online teaching most?

First, we focused specifically on OR as a key subject of study at the Faculty of Materials Science and Technology in Trnava, the Slovak University of Technology in Bratislava, Slovakia (further MTF STU). The findings from online surveys (years 2015-2019) about the quality of the educational process at MTF STU with a focus on OR were revised.

Another important source of research was secondary data for comparing OR areas of OR and the use of software in teaching OR at other universities in the Czech and Slovak Republic. Therefore, several online questionnaire surveys and also a semi-structured interview were used in the data collection.

The third source was focused on the education of students during the crisis in 2020-2022 (COVID), with a sample of bachelor's and master students studying at the two biggest and most important institutions of higher technical education in Slovakia – the Slovak University of Technology in Bratislava (STU) and the Technical University in Košice (TUKE).

4 Teaching Operation Research at Universities in Slovakia and the Czech Republic

The results of surveys considering the subject OR at universities in Slovakia and in Czech were analyzed, and the interconnections using both primary and secondary data. Not only the Slovak but also the Czech Republic has been chosen due to the common history of education in the years 1918-1991, linguistic proximity, and continuing current scientific and pedagogical co-operation.

4.1 Operation Research at Universities in Slovakia

Analysis of current state in teaching of the subject Operational Research in tertiary education at Slovak universities was carried out using the questionnaire method in

year 2015. As a supporting tool of the analysis university websites was carried out to determine the level of teaching of the subject OR (or related subjects) at universities in Slovakia (mentioned method was implemented also in year 2019). The survey focused on economic and technical higher education institutions, therefore higher education institutions with an artistic, medical, social focus were excluded from the sample, and foreign higher education institutions were not approached either. Using introduced pattern 12 public universities, 2 state HEIs and 4 private universities were addressed by the questionnaire. 14 completed questionnaires were returned, which represents a 77% return of questionnaires. To compare teaching issues, 50 OR areas were selected for the questionnaire, covering the areas of linear, nonlinear, and dynamic programming, decision theory, game theory, stochastic and simulation models, graph theory and project management methods, inventory models, collective service, recovery, and some other associated areas. Based on the questionnaire and information on university websites, we find out the number of subjects related to the monitored area, if subjects are compulsory for bachelor or master studies, and especially the educational content of subjects. The scope of teaching OR at some universities may be larger than indicated in the survey, as some areas of OR extend into other disciplines (mathematics, statistics, logistics, project management, etc.). Next, results from the questionnaire survey relevant to the focus of the article are presented.

The results stated that OR is taught at a doctoral, master and bachelor degree as follows: 64% of the respondents answered that it is taught at a master's degree, 36% at a bachelor's degree and 21% at third (doctoral) degree.

The key areas taught from OR are linear programming (LP) (79%); then the formulation of models and the graphical solution of LP problems (64%); problem-solving using software is mentioned by 57% of respondents and 50% of respondents mentioned numerical solution of LP problems, the duality of LP tasks, traffic problem and assignment problem. The least taught are the following areas: parametric and quadratic programming, Markov chains, neural networks, multicriteria and target programming, group decision making, and production planning (Fig. 2). We assume that areas that are less represented in the subject of OR are taught at universities in other subjects, such as logistics, production management, project management, statistical methods, and others.

In another question, 14% of respondents confirmed the interconnection of the OR issues with other subjects such as game theory, logistics, project management, systems to support mathematical modeling, quantitative methods in economics, methods, and models of efficiency evaluation, optimization. The key areas of OR taught in other subjects at universities are network analysis methods and LP tasks.

As software used in OR, 64% of respondents reported Microsoft Excel.

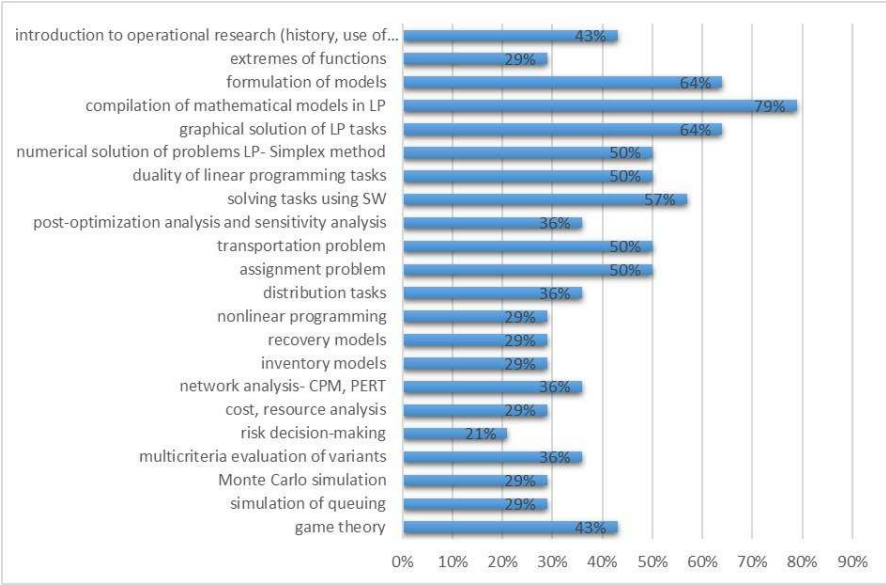


Figure 2
Areas taught within Operation research subject (drawn by the authors)

Online resources have been identified as the main study material for OR. It is positive for students, mainly because the study material is available 24 hours a day 7 days a week, and in every place where there is the Internet.

4.2 Operation Research at Universities in Czechia

In the Czech Republic, a survey was conducted in the previous period (2008, 2010) to compare the teaching of the OR subject and subsequently a comparison of the teaching at universities in the Czech Republic using methods of multicriteria evaluation of variants [26, 27] (The conducted results of the survey were revised and confirmed in years 2015, 2018 and 2019 using surveys in the final theses under the supervision of Kuncová [28, 29, 30], who was also involved in the realization of previous survey in 2008 and 2010). The survey was conducted on a selected sample of 18 public universities and was carried out based on data presented on the websites of individual universities. According to research 81 topics were gradually selected for comparison of teaching (covering the areas of linear, nonlinear, integer, quadratic and dynamic programming, decision theory, game theory, etc.) At the end of the comparison, it was found that 11 areas (out of 81 monitored) are taught at all compared universities:

introduction - practical use of models, formulation of models, LP problems - mathematical models, solution options, graphic design, duality, numerical solution

- simplex, software solution, post-optimization analysis, traffic problems (tasks), models of collective service, simulation - Monte Carlo method.

Based on the results of the questionnaire survey at universities in Slovakia and the results of the survey in the Czech Republic, we concluded that 11 areas (out of 81 monitored in the Czech Republic and out of 50 monitored in the Slovak Republic) are taught at all compared universities.

Also, software support for teaching the subject OR is comparatively provided at universities in the Slovak Republic and in the Czech Republic, the most frequently used are Microsoft Excel, Microsoft Project, LINDO, LINGO, WITNESS. Surveys also show that OR methods are taught in both economics and technical fields and most often are used for online learning e-scripts, e-learning platforms, PowerPoint presentations, e-consultations, etc.

5 Comparison of Teaching OR Before and During Covid Crisis

The authors compared the teaching of the OR subject before and during the Covid-19 crisis. The subject OR is taught at 2nd year of Master study at MTF STU in the winter term for following study programs: industrial engineering and management; personal policy, production technologies and production management; process automation and informatization in industry. In the standard (contact) way of teaching, students had lectures and exercises in the range of four hours per week, while they could have an appointment for face-to-face or via email consultations. In the standard situation, students can borrow literature for free (textbooks, lecture books, and books) in the academic library and/or study it in study room of the MTF STU during its opening hours or use the online textbook "Operation research" by authors H. H. Chovanová, P. Sakál, A. Štrpka and others. Students have three textbooks with lectures (scripts), two of them are available as well as e-scripts, and three scripts with instructions for practicing calculations, which are also available as e-scripts.

The survey about the quality of the educational process used for identification of the needs of students was analyzed from time period 2015-2019. The average number of students studying OR is about 319 students. Participation in the survey varies between 2% and 16% of the enrolled students.

Recommendations for improvement were as follows:

- Need to calculate more examples within key areas of OR;
- The time intensity of the curriculum (cannot be adapted to the individuality of the student);

- Dislike of the manual calculation in OR;
- "Lack" or not enough possibilities of calculation using software (SW: POM QM for Windows is used within the subject, and it is also possible to use MS Excel adds Solver).

During the Covid-19 pandemic, teaching conditions changed significantly, changing toward distance learning through GSuite and/or Microsoft Teams. Lectures and calculations are conducted only online using an interactive whiteboard - eBeams Edge, while lecture is recorded, and recordings are freely provided to students as mp4 files.

Students using GSuite and/or Microsoft Teams can be active and respond to the teacher's questions or ask questions to the teacher. Among the identified shortcomings of online teaching, students included technical problems in using online application (audio, video dropouts and connection difficulties). According to semi structured interviews with students the following positives were found during the online teaching as follows: the possibility to view the lesson more than once from the recording (the student can adapt the speed of the downloaded curriculum to his / her individual needs), the teacher has the opportunity to record more comprehensive videos and thus explain more in detail. Some students have stated as an advantage that they can now to follow the lesson for longer and without interruption; thus understanding is better.

If we summarize what contact and distance education also have similarities, because all lectures are provided as ppt files via the Academic Information System at MTF STU (AIS), students have access to the "e-library", where they have e-scripts available 7 days a week and 24 hours a day, subject-relevant online files that help to master the curriculum are uploaded to AIS as well as in clouds (MSTEAMS). It is possible to consult with teacher online using TEAMS or Gsuite platform or via e-mail.

6 Teaching During Crisis -Survey at STU and TUKE

The Student Council of Higher Education and the Slovak Accreditation Agency for Higher Education conducted a comprehensive analysis of education during the pandemic situation in 2020 through a joint questionnaire survey [31].

The joint questionnaire survey of the Student Council of Universities and the Slovak Accreditation Agency for Higbriher Education took place from 10th to 22nd June, 2020. The online questionnaire was placed on the websites of both analyzed institutions and was disseminated using social networks and official communication with individual universities. In total, 3490 respondents from 23 universities, representing 105 different faculties, answered the questions.

The analyzed data about education of students during the crisis in 2020 (COVID), include a sample of bachelor's and master students studying at the two largest and most important institutions of higher technical education in Slovakia – the Slovak University of Technology in Bratislava (STU) and the Technical University in Košice (TUKE). The article will only focus on the results of the two already mentioned universities with a technical focus - Slovak University of Technology in Bratislava (STU) and Technical University in Košice (TUKE). From the respondents participating in the survey studied 892 at STU and 612 at TUKE. The sample obtained, which represents both genders, each faculty and several study programs, correspond to the population.

When comparing the satisfaction, awareness and comprehension of online education at technical universities STU and TUKE, we can state that there exist many similarities in exanimated areas. Within used five - point Likert scale (strongly agree, mostly agree, not sure, mostly agree and strongly disagree) students confirm that they are satisfied with online education in the study program with 61.92% at STU (strongly agree 26.91% and mostly agree 28.8%) and with 57.58% at TUKE (from them 32.68% strongly agree). Only 8.9% of students strongly disagree with online education satisfaction at STU, and 13.06% strongly disagree at TUKE. Mostly disagree with this online form 14.69% at STU, 12.42% mostly disagree at TUKE. Regarding awareness, most of students agreed strongly or mostly to be informed enough to handle distance (online) learning namely almost 62% of STU students and 63.57% of TUKE students. The results of comprehensiveness of online lessons compared to face-to-face teaching show that 16.48% fully agree and 31.17% partially agree with STU, while 14.22% strongly agree with TUKE and 19.12% partially agree. Dissatisfaction with the comprehensiveness of online teaching compared to face-to-face teaching was indicated by 15.47% completely and mostly 24.44% at STU, this is also confirmed by the results at TUKE where 25.65% mostly disagree and 30.23% of survey respondents strongly disagree.

Important finding in area of technical higher education is that online education is regarding respondents not comprehensive 15.47%, mostly not comprehensive with 24.44% from STU and confirmed by respondents from TUKE strongly not comprehensive 32.97% and mostly not comprehensive 24.33% (Fig. 3, Fig. 4).

In second part of the survey was analyzed which methods of online teaching were used the most at STU and TUKE, (online lectures, video conferences, e-learning, shared presentations, work on seminar papers and self-study of scripts and study literature). The most preferred were the presentation shared by the teacher (STU 23.65%, TUKE 27.94%), the study of literature (STU 23.21%, TUKE 37.75%) and least used were claimed from archive (not in real time) 8.30% at STU and the lectures online 8.72% at TUKE.

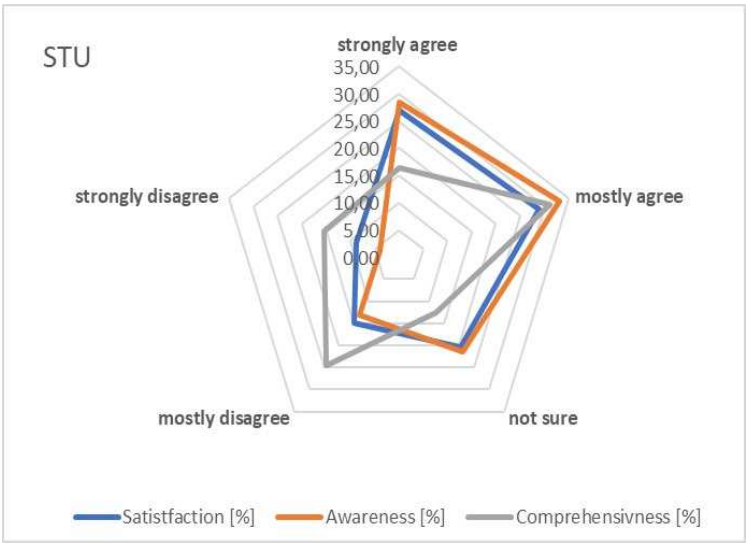


Figure 3
The data visualization - comparing satisfaction, awareness and comprehensiveness of online education at STU (drawn by authors)

Third, after visualization and comparing satisfaction, awareness and comprehensiveness in radar chart for both analyzed institutions (Fig. 3, Fig. 4), it indicates the relationship between awareness and their satisfaction with online teaching. Therefore, it is very important to inform and make familiar students with new possibilities in education process and their specifics, and so to increase their awareness and interest to be satisfied.

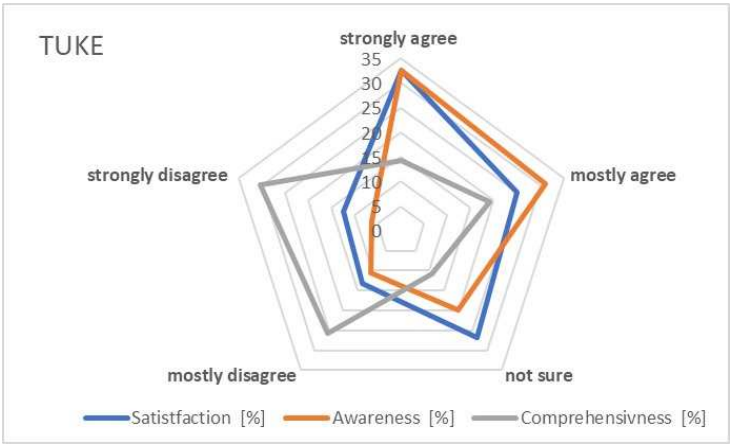


Figure 4
The data visualization - comparing satisfaction, awareness and comprehensiveness of online education at TUKE (drawn by authors)

Conclusions

Following the results and materials that the readiness of universities for the distance form of teaching we can conclude that teaching at the beginning of the pandemic was chaotic and new challenge for both students and teachers, but in a short time most of them were able to adapt to the situation so that the quality of the teaching process for students was secured. Education is a process during which new knowledge, skills and attitudes are acquired and developed. [32] The need for better technical support for teachers and education also emerged. E-learning proved to be a powerful tool for education, it made it possible to continue the educational process. Digital transformation of education is required in tertiary education to make the learning more attractive and easier for students. [33]

This also confirmed that information communication technology is among the latest innovations that have revolutionized various operations in the world [6, 34], including education and enables the enhancement of learning in the online environment [35].

On the other hand, based on survey results, it is confirmed from the students' point of view that E-learning methods cannot fully replace proper face to face education process, and to bring advantages of both it is best to use combination of both methods to achieve optimal results and improvement. In our opinion, the new crisis experience will help in the modernization of teaching and learning at universities especially in technical education. We agree with [4] that it is particularly important ICT in the field of education since it has recently created such platforms and opportunities that have facilitated to some extent the acquisition of knowledge.

An interesting trend is emerging as the possibility of switching to a combined teaching method with the use of online methods and modern information technologies.

During the first wave of the pandemic (March 2020), the new research project KEGA *"The implementation of innovative educational methods and MM guide for decision making area and application of analytical methods in the teaching process of selected subjects in the field of Industrial engineering"* under the Slovak ministry of education was submitted by Institute of Industrial Engineering and Management at MTF STU. The project was approved in January 2021. The main goal of the project is the creation of a portal for e-learning, a pilot for the three core subjects OR, Statistical methods and Business logistics, while the portal will contain, among other things, presentations of the core subjects of the subjects, solved examples (also with video support), electronic scripts, etc. As part of research project results, a survey is planned for students to specify, how the content and e-learning portal structure helped to acquire new knowledge and skills in mentioned areas. The new study will comprise and analyze the results of students before, during, and after the end of the pandemic including the academic years 2022/2023 and 2023/2024. In the future, the portal will represent support for

students' self-study and also an opportunity for professional and lay faithful to learn about the methods, procedures and models that the given subjects offer. In the future, it is planned that other subjects taught will be gradually added to the portal. The recent past and current situation confirms the trend that universities will have to be prepared for similar circumstances, as was and still is the Covid-19 pandemic. Therefore, further research will analyze the possibility of creating virtual laboratories and supporting education using the tools of virtual /augmented/mixed reality. The main aim of further research will be to analyze the effectiveness of implementing innovative methods o in higher education, including: Using virtual reality technology, Blended learning, Project-based learning, Cloud computing teaching, Flipped Classroom.

Acknowledgement

This paper was written with the financial support of the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic as a part of the project KEGA No.021STU-4/2021- "*The implementation of innovative educational methods and MM guide for decision making area and application of analytical methods in the teaching process of selected subjects in the field of Industrial engineering.*"

References

[1] L. Jurík, P. Sakál: Competencies of Managers, as part of the Intellectual Capital in Industrial Enterprises. ECIC 2014: proceedings of the 6th European Conference on Intellectual Capital, Trnava, 2014, pp. 368-376

[2] E. L. P. Montero, J. B. R. Zarta: Operations Research: from Strategy to Meaningful Learning, TECCIENCIA, 14(26), 2019, Available online: http://dx.doi.org/10.18180/tecciencia.2019.26.2

[3] J. S. Fu: ICT in Education: A Critical Literature Review and Its Implications, International Journal of Education and Development using Information and Communication Technology, Vol. 9, Issue 1, 2013, pp. 112-125, Available online: https://files.eric.ed.gov/fulltext/EJ1182651.pdf

[4] W. Sh. Basri, J. A. Alandejani, F. M. Almadani: ICT Adoption Impact on Students' Academic Performance: Evidence from Saudi Universities, Education Research International, Vol. 2018, Available online: https://doi.org/10.1155/2018/1240197

[5] J. E. Lawrence: Examining the factors that influence ICT adoption in SMEs: a research preliminary findings, International Journal of Technology Diffusion (IJTD), Vol. 6, No. 4, 2015, pp. 40-57

[6] R. Oliver: The Role of ICT in Higher Education for the 21st Century: ICT as a Change Agent for Education, Curtin, 2002, Available online: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.9509&rep=rep1&type=pdf

[7] ICT (information and communications technology, or technologies), Available online: https://searchcio.techtarget.com/definition/ICT-information-and-communications-technology-or-technologies

[8] B. Nagy, M.n Váraljai, A. Mihalovicsné Kollár: E-learning Spaces to Empower Students Collaborative Work Serving Individual Goals, Acta Polytechnica Hungarica, 17(2), 2020, Available online: http://acta.uni-obuda.hu/Nagy_Varaljai_Mihalovics-Kollar_99.pdf

[9] E. Mushtaha, S. A. Dabous, I. Alsyouf, A. Ahmed, N. R. Abdraboh: The challenges and opportunities of online learning and teaching at engineering and theoretical colleges during the pandemic, Ain Shams Engineering Journal, Volume 13, Issue 6, 2022, Available online: https://doi.org/10.1016/j.asej.2022.101770

[10] M. C. Hughes, B. W. Henry, M. R. Kushnick: Teaching During the Pandemic? An Opportunity to Enhance Curriculum, Pedagogy in Health Promotion: The Scholarship of Teaching and Learning, 2020, Vol. 6(4) 235-238, Available online: https://doi.org/10.1177/2373379920950)

[11] L. E. Kim, L. Oxley, K. Asbury: What makes a great teacher during a pandemic? Journal of Education for Teaching, 48:1, 129-131, 2022, DOI: 10.1080/02607476.2021.1988826

[12] E. Vázquez-Cano: Mobile distance learning with smartphones and apps in higher education, Education. Sci.: Theo. Pract., 14 (4) (2014) pp. 1505-1520

[13] M. Lynch, U. Kamovich, K. K. Longva, M. Steinert: Combining technology and entrepreneurial education through design thinking: Students' reflections on the learning process, Technological Forecasting and Social Change, Volume 164, 2021, Available online: https://doi.org/10.1016/j.techfore.2019.06.015

[14] Sir J. Daniel: Education and the COVID-19 pandemic, Prospects, 49, 2020, pp. 91-96, DOI: https://doi.org/10.1007/s11125-020-09464-3

[15] B. Alexander, K. Ashford-Rowe, N. Barajas-Murphy, G. Dobbin, J. Knott, M. McCormack, J. Pomerantz, R. Seilhamer, N. Weber: EDUCAUSE Horizon Report: 2019, Higher Education Edition, Louisville, CO (2019)

[16] G. L. Tortorella, G. Narayanamurthy, V. Sunder M., P. A. Cauchick-Miguel: Operations Management teaching practices and information technologies adoption in emerging economies during COVID-19 outbreak, Technological Forecasting and Social Change, Volume 171, 2021, Available online: https://doi.org/10.1016/j.techfore.2021.120996

[17] A. Ginige, H. Mayr, D. Plexousakis, V. Ermolayev, M. Nikitchenko, G. Zholtkevych, A. Spivakovskiy: Information and communication technologies in education, research, and industrial applications, ICTERI

2016, Cham: Springer Nature, 2017, Available online: https://link.springer.com/content/pdf/10.1007/978-3-319-69965-3.pdf

[18] (J. W. Redinger, P. B. Cornia, MD; T. J. Albert: Teaching During a Pandemic, J Grad Med Educ (2020) 12 (4): 403-405, Available online: https://doi.org/10.4300/JGME-D-20-00241.1)

[19] E. Mari et al.: Teaching during the Pandemic: A Comparison in Psychological Wellbeing among Smart Working Professions, Sustainability, 2021, Available online: https://doi.org/10.3390/su13094850

[20] T. Almpanis, P. Joseph-Richard: Lecturing from home: Exploring academics' experiences of remote teaching during a pandemic, International Journal of Educational Research Open, Volume 3, 2022, Available online: https://doi.org/10.1016/j.ijedro.2022.100133

[21] A. Parpala, S. Niinistö-Sivuranta: Leading Teaching during a Pandemic in Higher Education—A Case Study in a Finnish University. Educ. Sci. 2022, 12, 147, Available online: https://doi.org/10.3390/ educsci12030147

[22] J. Guncaga, J. Lopuchova, V. Ferdianova, M. Zacek, Y. Ashimov: Survey on Online Learning at Universities of Slovakia, Czech Republic and Kazakhstan during the COVID-19 Pandemic. Educ. Sci. 2022, 12, 458, Available online: https://doi.org/10.3390/ educsci12070458

[23] R. Companys, I. Ribas: "Some Trends and Applications of Operational Research/Management Science to Operations Management", International Journal of Production Management and Engineering, 2014, Available online: https://polipapers.upv.es/index.php/IJPME/article/view/3459/3763

[24] H. H. Chovanová: Uplatnenie vybraných metód operačnej analýzy v manažérskom rozhodovaní v podmienkach priemyselných podnikov (In English – Application of selected methods of operations analysis in managerial decision-making in the conditions of industrial enterprises) habilitation thesis, MTF STU, Trnava, 2018

[25] A. Raffaele, A. Gobbi: Teaching Operations Research Before University: A Focus on Grades 9-12, Operations Research Forum volume 2, Article number: 13, 2021, Available online: https://doi.org/10.1007/s43069-021-00054-3

[26] M. Kuncová, M. Lagová: Srovnání výuky a simulací na vysokých školách v ČR a SR (In English – Comparison of teaching and simulations at universities in the Czech Republic and Slovakia, ERIE 2008. Praha: ČZU PEF, 2008, pp. 107-115

[27] J. Kalčevová, M. Kuncová, B. Bokšteflová, J. Mašátová, P. Smrčka: Srovnání výuky operačního výzkumu na vysokých školách v České republice s využitím metod vícekriteriálního hodnocení variant (In English – Comparison of the Operational Research Tuition at the Universities in the Czech Republic by Mulcticriteria Decision Making Methods), Logos

Polytechnikos, 1(2), 2010, pp. 25-36, Available online: https://www.vspj.cz/tvurci-cinnost-a-projekty/casopisy-vspj/logos-polytechnikos, last accessed 2020/03/10

[28] Z. Ľochová: Výuka operačného výskumu na vysokých školách v ČR a SR – viackriteriálne hodnotenie (Tuition of the operational research at universities in the Czech Republic and Slovakia – multi-criteria evaluation), Praha: VŠE, 2019

[29] N. Sheshneva: Výuka operačního výzkumu na vojenských vysokých školách v EU – popis a srovnání (Tuition of the Operational Research at EU Military Universities - Description and Comparison), Praha: VŠE, 2018

[30] Š. Kubíková: Výuka operačního výzkumu na ekonomických vysokých školách v EU - popis a srovnání (Comparison of the Operational Research Tuition at the Universities in EU), Praha: VŠE, 2015

[31] M. Bílik: Survey results - Impact of COVID-19 on students, 2020, Available online: https://srvs.eu/2020/07/23/vysledky-prieskumu-dopad-covid-19-na-studentov/

[32] L. Lachvajderova, J. Kadarova., E. Sanchis., J. M. Vano. Significance of employee education for the development of the company. Acta Tecnología, Vol. 9, Issue 1, 2013, pp. 25-29, Available online: https://doi.org/10.22306/atec.v9i1.164

[33] M. Ingaldi, R. Ulewicz, D. Klimecka-Tatar, Creation of the university curriculum in the field of Industry 4.0 with the use of modern teaching instruments - Polish case study, Procedia Computer Science, Vol. 217, 2023, pp. 660-669, Available online: https://doi.org/10.1016/j.procs.2022.12.262

[34] M. Hilbert, P. López: The world's technological capacity to store, communicate, and compute information, Science, Vol. 332, No. 6025, 2011, pp. 60-65

[35] P. T. Mai, A. Tick: Cyber Security Awareness and Behavior of Youth in Smartphone Usage: A Comparative Study between University Students in Hungary and Vietnam, Acta Polytechnica Hungarica, 18(8), 2021, pp. 67-89, Available online: http://acta.uni-obuda.hu/Mai_Tick_115.pdf

Providing Guaranteed Performances for an Enhanced Cruise Control Using Robust LPV Method

Balázs Németh

Institute for Computer Science and Control (SZTAKI), Eötvös Loránd Research Network (ELKH), Kende u. 13-17, H-1111 Budapest, Hungary
e-mail: balazs.nemeth@sztaki.hu

Abstract: The paper proposes the design of an enhanced cruise control system for automated vehicles. The control strategy has three components, such as a predictive optimal control, a robust Linear Parameter-Varying (LPV) control and an optimization-based supervisor. In the design process of the control, primary performances (safety and speed limitation requirements) and secondary performances (economy and traveling time criteria) are considered. These performances are guaranteed through the different control components. The enhanced cruise control is able to provide solution to complex cruise control scenarios with a guaranteed performance level. The effectiveness of the method is illustrated through various simulation examples, in which the loss of the performance level is avoided by the proposed control.

Keywords: automated vehicle cruise control; LPV control design; performance specifications

1 Introduction and Motivation

Various safety, economy and comfort requirements against automated vehicles pose complex decision and control challenges for research teams in the field of vehicle control design. A possible solution to the adaptation to the environment of the vehicle is to use increased number of information on the road and traffic through vehicle to vehicle (V2V) and vehicle to infrastructure (V2I) communication. The information is used in different layers of the longitudinal control in automated vehicles, such as perception, navigation, design of the route and the speed profile [1].

In the recent years several design methodologies in the field of enhanced energy efficient driving systems on several vehicle control tasks have been developed. An overview about the principles of the energy efficient cruise control has been proposed in [2]. The consideration of forthcoming terrain characteristics has been handled by using a receding horizon control in real experiments in [3]. The work of [4] has presented a deep learning-based eco-driving solution to electric vehicles, in which information about the surrounding vehicles has also been incorporated. Eco-cruise control system for automated vehicles in intersection scenarios has been implemented

in [5]. Systematic design and analysis methods for predictive cruise control systems with the consideration of road and traffic information have been presented in [6].

Due to various external information sources the achieved performance level of the of automated vehicle control can depend on the quality of the communicated data [8]. The source of the information might be static datasets, e.g., terrain characteristics or speed limit rules on the road. Nevertheless, some important information can vary dynamically, such as variable speed limits on high-speed roads, traffic flow information and actual motion information of the surrounding vehicles [7]. Although these information can be important to provide energy-efficient and comfortable motion for automated vehicles, the degradations in the communicated data, and thus, a challenge is to build reliable architectures, which are less dependent on the degradation of the communicated data and thus, the predefined performances can be guaranteed.

The performances of enhanced cruise control systems can be classified based on their priorities. There are primary performance requirements in safety-critical systems, especially in cruise control systems, which must be guaranteed by the control during the entire operation of the closed-loop system. Primary performances are related to keeping safe distance, i.e. from the preceding vehicle and from the follower vehicle in the case of a lane change maneuver. Moreover, a primary performance is to keep vehicle speed in a bounded range of the speed limit, with which the violation of the speed regulations or the dangerously slow motion of the vehicle in a high-speed road can be avoided. Moreover, the performances of the cruise control systems may have another group, the secondary performances, which are requested to consider by the control system, e.g. comfort criteria, energy consumption minimization or traveling time requirements. These performances are requested to maintain due to the expectations of the users, but they can be violated in critical situations, e.g. if a collision is predicted. The presented various performance requirements demand increased number of information sources, especially communicated data.

The goal of the paper is to propose a design framework for enhanced cruise control systems, with which guarantees on the primary performances are provided. In the framework two controllers are designed. It is designed a controller based on the robust Linear Parameter-Varying (LPV) control theory, which uses on-board sensor information and limited number of external information. The controller is able to provide guaranteed primary performances together with a supervisory strategy. The minimum performance level of the enhanced cruise control on the primary performances is equivalent to the performance level of the robust LPV control on the primary performances. Furthermore, a predictive optimal cruise control is also designed, in which several external information is incorporated. The proposed predictive control system is able to maintain primary and secondary performances effectively, but the primary performances cannot be guaranteed for all scenarios. The design of the predictive cruise control is based on the method, which is presented in [6]. During the cruising of the automated vehicle both controllers compute their control signal parallel. The control intervention based on the two signals is computed by a supervisory strategy.

The enhanced cruise control for automated vehicles is composed of the robust LPV

control, the nonlinear predictive optimal control and the supervisor. Interconnection between them is created by scheduling variables and known uncertainties, which are taken part of the robust LPV control design. The motivation behind the robust LPV formalism is flexibility, which can be achieved by the adaptation capability of the controller through the selection of the scheduling variable.

The contribution of the paper is an enhanced cruise control system, which is able to solve the complex cruising problem with several performances for automated vehicles. The novelties of the proposed method are summarized as follows. First, the proposed method provides theoretical guarantees on the performance level of the primary performances. Second, the guaranteed performance level of the system is less dependent on the degradation of the communicated data from the external information sources, which can provide an improved level of safety for automated vehicles.

The paper is organized as follows. The strategy of the nonlinear predictive optimal control, which considers various information sources is presented in Section 2. The robust LPV-based framework for modeling and control design is proposed in Section 3. Section 4 proposes the supervisory strategy, and than, the design of the robust LPV controller in an iterative framework is presented in Section 5. Section 6 illustrates the effectiveness of the proposed method. Finally, the consequences of the design method are summarized in Conclusions.

2 Design of Predictive Cruise Control for Automated Vehicles

The role of the section is to present the design method of the predictive cruise control briefly. The aim of the description is to provide an overview about the formulation of the performances and the incorporation of the external information in the predictive control problem. A thorough discussion of the method is found in [6].

The predictive cruise control can use various information sources, i.e., in this paper four different information sources are considered to be available. First, the automated vehicle has information from topography database, which provides altitude and road curvature information. The road section ahead of the vehicle is divided into n number of segments, where the lengths of the segments are selected to have constant inclinations. Second, the vehicle has information about speed limitations on the road segments. Since speed limitations can also depend on the actual road construction works and variable speed limit signs in high-speed roads, it can require information from static road map and V2I communication. Third, information about the average traffic speed on the forthcoming road section and the state of the traffic lights expect communication with the traffic control system. Fourth, information about the actual speed and the positions of the surrounding vehicles can require V2V communication and on-board sensors, e.g. radar measurements.

The performances of the predictive cruise control are formed as follows. A primary performance of the vehicle is to keep safe distance from the preceding vehicles in the own lane and from the follower vehicles in the case of a lane change maneuver on the entire horizon. As an assumption, it is considered that the vehicles move in the

same directions on the road. Moreover, motion information about the surrounding vehicles are considered in a predefined region of interest, which leads to N_p number of preceding vehicles and N_f number of follower vehicles. Formally, it leads to the conditions

$$e^{k_p} + \sum_{i=1}^{j} \left(\eta_i^{k_p} - \xi_i \right) \geq d_{safe}, \quad \forall j \in \{1,...,n\}, \quad \forall k_p \in \{1,...,N_p\} \tag{1a}$$

$$e^{k_f} + \sum_{i=1}^{j} \left(\xi_i - \eta_i^{k_f} \right) \geq d_{safe}, \quad \forall j \in \{1,...,n\}, \quad \forall k_f \in \{1,...,N_f\} \tag{1b}$$

where k_p, k_f represent the indexes of the preceding and follower vehicles, e^{k_p}, e^{k_f} are the actual distance between vehicle k_p, k_f and the automated vehicle and d_{safe} is the requested safe distance. Index j represents the road segment and $\sum_{i=1}^{j} \xi_i$ is the predicted longitudinal displacement of the automated vehicle until step j and $\sum_{i=1}^{j} \eta_i^{k_p}$, $\sum_{i=1}^{j} \eta_i^{k_f}$ are the predicted displacements of vehicle k_p and k_f. Relations in (1) represent that the predicted distance between the automated vehicle and a surrounding vehicle until horizon j cannot be smaller then the predefined safe distance. If the relations are guaranteed, the safe distances from all surrounding vehicles on the entire horizon are kept.

Further primary performance of the control is keeping vehicle speed in a limited range around reference speed $v_{ref,i}$ in segment i. $v_{ref,i}$ is selected based on the speed limitation, road curvature, average traffic speed [6, 9]. The performance is formed as

$$\dot{\xi}_i \in \left[v_{min,i}; v_{max,i} \right], \quad \forall i \in \{1,...,n\}, \tag{2}$$

where $\dot{\xi}$ is the speed of the automated vehicle and $v_{min,i}, v_{max,i}$ values are the limits (minimum and maximum) of the speed range, in which the vehicle speed can vary. Performance (2) guarantees keeping speed limitations. Furthermore, it guarantees the avoidance of the dangerously slow motion of the automated vehicle. The values of $v_{min,i}, v_{max,i}$ are derived from the value of the speed reference $v_{ref,i}$ on each segment, e.g. $-20\%, +5\%$ related to $v_{ref,i}$.

One of the most important secondary performance in the cruise control problem is to achieve minimum control intervention on the road horizon ahead of the automated vehicle, which leads to the criterion

$$\sum_{i=1}^{n} |F_{l,i}| \rightarrow min, \tag{3}$$

where $F_{l,i}$ represents traction/braking force on segment i of the horizon.

Another secondary performance is to minimize traveling time of the vehicle. Since the shortest traveling time is equivalent with the maximum speed motion of the vehicle, it can be transformed to the speed objective as

$$|v_{max,i} - \dot{\xi}_i| \rightarrow min, \quad \forall i \in \{1,...,n\}. \tag{4}$$

The motions of the automated vehicles can have impact on the characteristics of the traffic flow, because the speed profiles of the automated vehicles can differ from the speed profiles of the human-driven vehicles. This impact has increasing importance through the increase of the traffic density and the ratio of the automated vehicles in the traffic flow. A further secondary performance of the control is that motion of the automated vehicles must have advantageous impact on the traffic flow. It means that the output flow of the traffic network q_{out} must be maximized, such as

$$q_{out} \rightarrow max. \tag{5}$$

The relationship between q_{out}, the speed selection strategy of the automated vehicles, the ratio of the automated vehicles and the traffic density is characterized in [6].

Formulation of the Optimization Process

The computation of the actual control input $F_{l,1}$ is based on a predictive optimal control strategy, which considers the previously defined performance specifications [6]. It leads to a hierarchical optimization structure. In the low level of the structure a solution to the secondary performance problem is found, while in the high level the priority performance criteria are incorporated. The interconnection between the levels is provided by a parameter R, which is interpreted below.

In the low level of the optimization, weights Q, γ_i, $i \in \{1,..,n\}$ are defined to all the segments on the horizon. Their role is to define the importance of each segments in the design of the current speed. Weight Q determines the tracking requirement of the predefined actual reference speed $v_{ref,0}$, which is related to the current segment of the vehicle. The road inclinations α_i and the reference speeds on the segments of the horizon ahead of the vehicle are considered through the weights γ_i. The result of the predictive control is a reference speed λ for the vehicle, which is characterized by the weights Q, γ_i, such as

$$\lambda = \sqrt{\vartheta - 2s_1(1-Q)(\ddot{\xi}_0 + gsin\alpha)}, \tag{6}$$

where $\ddot{\xi}_0$ is the longitudinal acceleration, s_1 is the length of first road segment and ϑ incorporates force and reference speed components of the forthcoming road sections:

$$\vartheta = Qv_{ref,0}^2 + \sum_{i=1}^{n} \gamma_i v_{ref,i}^2 + \frac{2}{m} \sum_{i=1}^{n} s_i F_{di,r} \sum_{j=i}^{n} \gamma_j, \tag{7}$$

where the known longitudinal force resistance $F_{di,r}$ contains the road inclination in segment i.

The selection of Q, γ_i values are based on the secondary performance criteria. Actual control force can be expressed in a form, which depends on Q and γ_i. Through the transformation of (3) to a quadratic criterion $F_{l,1}^2 \rightarrow \min$, the following optimization problem is yielded

$$\left(\beta_0(\bar{Q}) + \beta_1(\bar{Q})\bar{\gamma}_1 + \ldots + \beta_n(\bar{Q})\bar{\gamma}_n\right)^2 \rightarrow \min \tag{8}$$

with the constrains $0 \leq \bar{Q}, \bar{\gamma}_i \leq 1$ and $\bar{Q} + \sum \bar{\gamma}_i = 1$. $\bar{Q}, \bar{\gamma}_i$ are the solutions of (8) and $\beta_0(\bar{Q}), \beta_i(\bar{Q})$ are matrices. Objective (8) is nonlinear in Q, but for a fixed Q value it leads to a quadratic optimization problem in γ_i with constraints. Secondary performance (4) is transformed to the minimization of the difference between the current speed and the reference speed, such as

$$|v_{ref,0} - \dot{\xi}_0| \rightarrow \min. \tag{9}$$

The solution of (9) is achieved by selecting the weights $\breve{Q} = 1$ and $\breve{\gamma}_i = 0, i \in [1,n]$, because in this case the automated vehicle tracks the actual reference speed value. The balance between the secondary performances is created through the selection of parameter $0 \leq R \leq R_{max} \leq 1$, such as

$$Q = R\bar{Q} + (1-R)\breve{Q} = 1 - R(1-\bar{Q}) \tag{10a}$$

$$\gamma_i = R\bar{\gamma}_i + (1-R)\breve{\gamma}_i = R\bar{\gamma}_i, \quad i \in \{1,..,n\}. \tag{10b}$$

If R is selected for a high value, the minimization of the control force of the automated vehicle is preferred. It can lead to a reduced speed for the vehicle, which considers the forthcoming road and traffic information in the computation of F_{l1}. But, if R has a low value, the speed of the vehicle is close to $v_{ref,0}$, which means that minimum traveling time is preferred. Thus, the selection of R has a high impact on the balance between the performances traveling time and control force, and consequently, on the speed profile of the automated vehicle. The value of R_{max} is determined by performance (5), which is related to the maximization of the traffic flow. The actual value of R_{max}, whose selection can result in high q_{out}, depends on the actual traffic density and the ratio of the automated vehicles in the traffic. The relationship is characterized by a nonlinear function, which is based on scenario-based studies [6].

In the high level of the optimization architecture the goal is to calculate R, with which the primary performances can be considered. The purpose of the optimization is to maximize R, which leads to an energy-efficient motion with advantageous impact on the traffic flow. Nevertheless, in the high level optimization the primary performances (1)-(2) are handled as constraints of the optimization process. Thus, R must be selected as high as possible, but the resulted speed profile must guarantee primary performances. The resulted high-level optimization problem is

$$\max_{[0,R_{max}]} R \tag{11}$$

such that the constraints (1)-(2) are guaranteed. The result of the optimization on the high-level is R, which is used in the computation of Q, γ_i, see (10). Furthermore, Q, γ_i are applied in (6), which induces a speed tracking problem, whose result is the actual control force $F_{l,1}$. Through the values of ξ_i and $\dot{\xi}_i$ in the constraints (1)-(2), the result of the low-level optimization has an impact on the high-level. Thus, the maximization of R is an iterative process, until the appropriate value is achieved.

The presented optimization process can provide excellent control force for the vehicle, which considers several performance requirements. However, it is difficult to verify the result of the optimization through in theory due to the following limitations of the method.

- The control strategy requires several communicated data, whose safety and security challenges have been presented in Section 1.
- The maintenance of several performances requires complex control structure, which contains a hierarchical nonlinear optimization process. Moreover, the optimization problem depends on the actual traffic scenario, e.g., the number of constraints depends on the actual number of vehicles. Thus, it is difficult to find a compact, offline solution of the optimization process, which can be examined, e.g., from the aspect of the parameter sensitivity. Instead, the method can be verified through simulations and experimental scenarios.
- Constraints (1)-(2) depend on the prediction of the preceding/follower vehicle motion. It presupposes a vehicle motion model for the vehicles in the local surroundings, whose difference from the real vehicle motion can degrade the primary performance.

Therefore, the resulted nonlinear predictive optimal control strategy cannot be used alone in the automated vehicle cruise control system. It is requested to find a control strategy, with which the primary performances can be guaranteed, while the advantages of the predictive cruise control can be preserved in most of the vehicle cruising. It leads to the concept of the enhanced cruise control, as proposed in the rest of the paper.

3 Architecture of the Enhanced Cruise Control System

The basic idea of the control strategy is to design a robust LPV controller and a supervisory strategy, which can modify the control input of the predictive cruise control if the primary performances are violated.

The output of the predictive cruise control is represented as

$$u_P = \mathscr{F}(y_P) \tag{12}$$

where $u_P = F_{l1}$ denotes the control input of the predictive cruise control, y_P vector contains the inputs of the controller with m_P elements and $\mathscr{F}$ represents the predictive cruise controller itself. Moreover, the control signal u_K is the output of a robust LPV controller, such as

$$u_K = \mathscr{K}(\rho_P, y_K) \tag{13}$$

where $\mathscr{K}$ represents the robust LPV controller and y_K is the vector of the measured signals with m_K elements. In (13) $\rho_P \in \rho_P$ vector contains the scheduling variable of the controller, which is derived from the following control rule.

The most important assumption of the proposed method is that the actual value of the control signal u can be expressed in a linear form of u_K. If the primary performances are not violated by u_P, then $u = u_P$ is selected. Thus, under the consideration that the primary performances are not violated, the relationship between u_K and u_P with the conditions is formed as

$$u_P = \rho_P^* u_K + \Delta_P^*, \quad \text{if} \quad \rho_P^* \in \rho_P, \quad \Delta_P^* \in \Lambda_P, \tag{14}$$

where ρ_P^* and Δ_P^* are time-dependent weighting signals. $\rho_P^* = [\rho_{P,min}; \rho_{P,max}]$, $\Lambda_P^* = [\Delta_{P,min}; \Delta_{P,max}]$ represent domains in (14), where $\rho_{P,min}$, $\rho_{P,max}$, $\Delta_{P,min}$, $\Delta_{P,max}$ are scalars. The sets of the domains are denoted by ρ_P, Λ_P. In (14) the conditions are guaranteed, if the primary performances are not violated. Thus, it can be find ρ_P^*, Δ_P^* pair and $u = u_P$. But, if $\rho_P^* \notin \rho_P$ or $\Delta_P^* \notin \Lambda_P$, the variables ρ_P^*, Δ_P^* are limited with the boundaries of ρ_P and Λ_P during the computation of the control signal u. In this case u can significantly differ from u_P. The previous cases lead to a control strategy, which contains all scenarios as

$$u = \rho_P u_K + \Delta_P, \tag{15}$$

where

$$\rho_P = \min\left(\max\left(\rho_P^*; \rho_{P,max}\right); \rho_{P,min} \right), \tag{16a}$$

$$\Delta_P = \min\left(\max\left(\Delta_P^*; \Delta_{P,min}\right); \Delta_{P,max} \right). \tag{16b}$$

Thus, the concept of providing guaranteed primary performances is based on the bounding of ρ_P, Δ_P, the relations in (16) result in $\rho_P \in \rho_P$ and $\Delta_P \in \Lambda_P$. Therefore, if it is possible to design a robust LPV control with the scheduling variable ρ_P, Δ_P and the uncertainty $\Delta_P \in \Lambda_P$, the primary performances can be guaranteed through the appropriate selection of ρ_P, Δ_P values. In the control architecture the supervisor is responsible for the selection of ρ_P, Δ_P values.

The architecture of the proposed enhanced cruise control strategy is shown in Figure 1. In the control process u_P and u_K are computed simultaneously. The role of the supervisor is to select ρ_P, Δ_P and to generate u based on the rule (15).

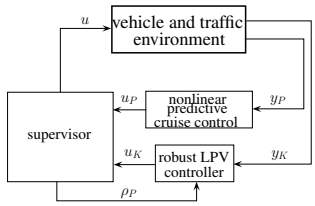


Figure 1
Scheme of the enhanced cruise control strategy

In the proposed enhanced cruise control architecture the selections of ρ_P, Δ_P and ρ_P, Λ_P have high influence on the operation of the system. If the ranges of the domains are selected small, u_P is often saturated due to the limitations of domain boundaries (16). But, if the ranges ρ_P, Λ_P have insufficiently high values, the resulted robust LPV controller can be conservative because of the increased robustness requirements. The objective of ρ_P, Δ_P, ρ_P, Λ_P selections is to provide u, with which u_P is approximated, while the primary performances are guaranteed. The objective results in the maintenance of the secondary performances, when u approximates u_P.

It might be suggested to find a joint control design and supervisor design algorithm, whose result is the selection of ρ_P, Δ_P and ρ_P, Λ_P simultaneously and thus, the optimization problem

$$\min_{\rho_P,\Delta_P,\rho_P,\Lambda_P} (u - u_P)^2 \tag{17}$$

is performed. However, the design of the robust LPV control is an offline process, which requires the preliminary selection of ρ_P, Λ_P, while the selection of ρ_P, Δ_P is influenced by the actual control interventions. Instead of the joint design (17) an approximation of the optimization is presented, in which the design of the robust LPV controller with the selection of ρ_P, Λ_P is divided from the design of the supervisor with the selection of ρ_P, Δ_P. The objectives of both design processes are the same, i.e. the minimization of the difference between u and u_P. Although the design processes are separated, both of them influence the achieved performance level of the enhanced cruise control.

4 Design of the Supervisory Strategy

The purpose of the supervisor is to provide control force actuation u through the selection of ρ_P, Δ_P, with which the primary performances during the cruising of the vehicle are guaranteed. The selection is based on the signals u_P, u_K, which are the inputs of the supervisor. The output of the supervisor u is constructed through (15), and moreover, the resulted ρ_P is used in the operation of the robust LPV control.

The design of the supervisor is based on the simplified longitudinal model of the vehicle:

$$m\ddot{\xi} = F_l - F_d, \tag{18}$$

where m is the mass of the vehicle. The state vector is $x = \begin{bmatrix} \dot{\xi} & \xi \end{bmatrix}^T$, where ξ represents the longitudinal motion of the vehicle, $w = F_d$ contains the longitudinal disturbances and $u = F_l$ involves the longitudinal control force. The state-space representation of the system is formed as

$$\dot{x} = Ax + \hat{B}_1 w + \hat{B}_2 u, \tag{19}$$

where x represents the state vector and $A, \hat{B}_1, \hat{B}_2$ are matrices in the representation of the system.

The state-space representation of the system is reformulated using the predefined control strategy (15), the control input of the robust LPV controller u_K is used in the expression $u = \rho_P u_K + \Delta_P$. Therefore, the state-space representation of the system (19) is reformulated through the relationship between u and u_K as

$$\dot{x} = Ax + B_1 w_K + B_2(\rho_P) u_K, \tag{20}$$

where the disturbance vector w_K in the state-space representation (20) is composed as $w_K = \begin{bmatrix} w & \Delta_P \end{bmatrix}^T$ and the matrices are $B_1 = \begin{bmatrix} \hat{B}_1 & B_2 \end{bmatrix}$ and $B_2(\rho_P) = \hat{B}_2 \rho_P$. Thus, the system is transformed into a robust LPV representation.

Specification of Conditions to Provide Guarantees on Primary Performances

The conditions to provide primary performances through the supervisor are specified based on the derived system formulation (20).

Performance (1) in the supervisor design process is focused on keeping safe distance from the closest preceding vehicle and from the closest follower vehicle of another lane, which leads to $N_p = 1, N_f = 1$. The goal of this simplification is to use less communicated data in the computation of ρ_P, Δ_P. The selection of the closest vehicles is performed continuously during the operation of the supervisor based on on-board sensor measurements. If a vehicle in the region of interest of the sensors is not found (e.g., the lane of the automated is empty ahead or behind), a virtual vehicle is considered to be on the bound of the region.

The prediction of the forthcoming distance d^{k_p} between the preceding vehicle and the automated vehicle is formulated based on their accelerations. The time-dependent function of $\ddot{d}^{k_p}$ is based on (18) as

$$\ddot{d}^{k_p}(t) = \ddot{\eta}^{k_p}(t) - \ddot{\xi}(t) = \ddot{\eta}^{k_p}(t) - \frac{F_l(t)}{m} + \frac{F_d(t)}{m}. \tag{21}$$

Through the integration of (21) the forthcoming speed difference in time T can be derived as

$$\dot{d}^{k_p}(T) = \int_0^T \ddot{d}^{k_p}(t)dt = \int_0^T \left(\ddot{\eta}^{k_p}(t) - \frac{F_l(t)}{m} + \frac{F_d(t)}{m} \right) dt \tag{22}$$

The integration requires knowledge about the functions $\ddot{\eta}^{k_p}(t)$ and $F_d(t)$. But, it can be difficult to predict the forthcoming acceleration command of the preceding vehicle and the forthcoming road disturbances. In case of a safe control strategy, these functions are substituted by constant values, which are resulted by worst-case scenarios. It is considered that $a_{min} \leq \ddot{\eta}^{k_p}(t) + \frac{F_d(t)}{m}$, where a_{min} represents the worst case scenario, when the preceding vehicle has maximum deceleration and the road disturbance has minimum value. The value of a_{min} is a design parameter, which can be selected based on preliminary experimental results. Using a_{min}, (22) is computed as

$$\dot{d}^{k_p}(T) = a_{min}T - \frac{F_l}{m}T + \dot{d}^{k_p}(0) = a_{min}T - \frac{F_l}{m}T + \dot{\eta}^{k_p}(0) - \dot{\xi}(0), \tag{23}$$

where $\dot{d}^{k_p}(0) = \dot{\eta}(0) - \dot{\xi}(0)$ is the speed difference at time $t = 0$ and $F_l(t)$ is assumed to be constant between 0 and T. The predicted distance between the vehicles is resulted by the integration of (23), such as

$$d^{k_p}(T) = \frac{a_{min}T^2}{2} - \frac{F_l T^2}{2m} + \dot{\eta}^{k_p}(0)T - \dot{\xi}(0)T + e^{k_p}, \tag{24}$$

where e^{k_p} is the measured distance between the preceding vehicle and the automated vehicle in time $T = 0$. The prediction in (24) requires measurement of the actual

distance e^{k_p} and the relative speed between the automated vehicle and the preceding vehicle $\dot{\eta}^{k_p}(0) - \dot{\xi}(0)$, which can be performed through on-board sensors, e.g., radar.

Similarly, the predicted distance between the automated vehicle and the follower vehicle k_f can be derived from the second derivative of the distance between them as $\ddot{d}^{k_f}(t) = \ddot{\xi}(t) - \ddot{\eta}^{k_f}(t)$. The worst-case scenario is characterized by the acceleration a_{max} through the expression $\ddot{\eta}^{k_f}(t) + \frac{F_d(t)}{m} \leq a_{max}$, which leads to the predicted distance

$$d^{k_f}(T) = \frac{F_l T^2}{2m} - \frac{a_{max} T^2}{2} + \dot{\xi}(0)T - \dot{\eta}^{k_f}(0)T + e^{f_p}. \tag{25}$$

The formulation of the primary performances, which means that safe distances from the preceding vehicle and the follower vehicle must be kept, are written as inequalities

$$\frac{a_{min} T^2}{2} - \frac{F_l T^2}{2m} + \dot{\eta}^{k_p}(0)T - \dot{\xi}(0)T + e^{k_p} \geq d_{safe}, \tag{26a}$$

$$\frac{F_l T^2}{2m} - \frac{a_{max} T^2}{2} + \dot{\xi}(0)T - \dot{\eta}^{k_f}(0)T + e^{f_p} \geq d_{safe}. \tag{26b}$$

Since $u = F_l$ and $u = \rho_P u_K + \Delta_P$ (15), the inequalities (26) are rewritten as

$$\frac{a_{min} T^2}{2} - \frac{(\rho_P u_K + \Delta_P) T^2}{2m} + \dot{\eta}^{k_p}(0)T - \dot{\xi}(0)T + e^{k_p} \geq d_{safe}, \tag{27a}$$

$$-\frac{a_{max} T^2}{2} + \frac{(\rho_P u_K + \Delta_P) T^2}{2m} + \dot{\xi}(0)T - \dot{\eta}^{k_f}(0)T + e^{f_p} \geq d_{safe}. \tag{27b}$$

Thus, it is necessary to select ρ_P, Δ_P for given u_K to guarantee the inequalities (27), with which the primary performance of keeping safe distance is guaranteed.

Performance of keeping vehicle speed in a given speed range (2) is also based on the simplified motion model of the vehicle (18). The predicted speed of the vehicle in T is resulted through the integration of the acceleration $\ddot{\xi}$ as

$$\dot{\xi}(T) = \int_0^T \left(\frac{F_l}{m} - \frac{F_d}{m} \right) dt. \tag{28}$$

Similarly to the derived conditions of keeping safe distance, the worst-case scenario is considered as $|F_d| \leq F_{d,max}$, where $F_{d,max}$ is considered to be the upper bound of the unknown disturbance. If $F_d > 0$, which means that the disturbance has accelerating effect, (28) is transformed as $\frac{F_l T}{m} + \frac{F_{d,max} T}{m} + \dot{\xi}(0)$, where $\dot{\xi}(0)$ is the actual speed of the automated vehicle and F_l is considered to be constant. If $F_d < 0$, (28) results in $\frac{F_l T}{m} - \frac{F_{d,max} T}{m} + \dot{\xi}(0)$, which means that F_d decelerates the vehicle. The condition for keeping vehicle speed in the given range $\left[v_{min,0}; v_{max,0}\right]$ is formed as

$$\frac{(\rho_P u_K + \Delta_P) T}{m} + \frac{F_{d,max} T}{m} + \dot{\xi}(0) \leq v_{max,0}, \tag{29a}$$

$$\frac{(\rho_P u_K + \Delta_P) T}{m} - \frac{F_{d,max} T}{m} + \dot{\xi}(0) \geq v_{min,0}, \tag{29b}$$

in which relations $u = F_l$ is transformed to $\rho_P u_K + \Delta_P$ (15). Thus, it is necessary to select ρ_P, Δ_P, with which conditions in (29) together with (27) are guaranteed.

Optimization in the Supervisor Strategy

The selection strategy of ρ_P and Δ_P is based on the optimization, which is presented in (17). In the supervisory process ρ_P, Δ_P are selected during the operation of the enhanced cruise control system.

The objective of the supervisor is to provide a control input u, which is as close as possible to u_P:

$$\left(u - u_P\right)^2 \rightarrow min. \tag{30}$$

Through (30) the control force intervention of the enhanced cruise control system approximates the output signal of the predictive cruise control. Moreover, during the selection of ρ_P, Δ_P the criteria of (27) and (29) must be guaranteed and the constraints $\rho_P \in \rho_P$, $\Delta_P \in \Lambda_P$ must also be satisfied.

The objective (30) using (15) is rearranged to a quadratic form as

$$\begin{aligned}(u-u_P)^2 &= \begin{bmatrix}\rho_P \\ \Delta_P\end{bmatrix}^T \begin{bmatrix}u_K^2 & u_K \\ u_K & 1\end{bmatrix} \begin{bmatrix}\rho_P \\ \Delta_P\end{bmatrix} + \begin{bmatrix}-2u_P u_K \\ -2u_P\end{bmatrix}^T \begin{bmatrix}\rho_P \\ \Delta_P\end{bmatrix} + u_P^2 \\ &= \begin{bmatrix}\rho_P \\ \Delta_P\end{bmatrix}^T \beta \begin{bmatrix}\rho_P \\ \Delta_P\end{bmatrix} + \omega^T \begin{bmatrix}\rho_P \\ \Delta_P\end{bmatrix} + u_P^2,\end{aligned} \tag{31}$$

in which u_P^2 is independent from ρ_P, Δ_P and thus, it can be eliminated during the minimization process (30).

The strategy of the supervisor is to compute ρ_P, Δ_P during the operation of the cruise control. In each step the following constrained optimization problem must be solved, which is yielded from (31) and the constraints (27), (29) together with the bounds on ρ_P, Δ_P:

$$\min_{\rho_L, \Delta_L} \begin{bmatrix}\rho_P \\ \Delta_P\end{bmatrix}^T \beta \begin{bmatrix}\rho_P \\ \Delta_P\end{bmatrix} + \omega^T \begin{bmatrix}\rho_P \\ \Delta_P\end{bmatrix}, \quad \text{subject to} \tag{32a}$$

$$-\frac{(\rho_P u_K + \Delta_P)T^2}{2m} + \frac{a_{min}T^2}{2} + \dot{\eta}^{k_p}(0)T - \dot{\xi}(0)T + e^{k_p} \geq d_{safe}, \tag{32b}$$

$$\frac{(\rho_P u_K + \Delta_P)T^2}{2m} - \frac{a_{max}T^2}{2} + \dot{\xi}(0)T - \dot{\eta}^{k_f}(0)T + e^{f_p} \geq d_{safe}, \tag{32c}$$

$$\frac{(\rho_P u_K + \Delta_P)T}{m} + \frac{F_{d,max}T}{m} + \dot{\xi}(0) \leq v_{max,0} + S, \tag{32d}$$

$$\frac{(\rho_P u_K + \Delta_P)T}{m} - \frac{F_{d,max}T}{m} + \dot{\xi}(0) \geq v_{min,0} - S, \tag{32e}$$

$$\rho_P \in \rho_P, \quad \Delta_P \in \Lambda_P. \tag{32f}$$

In (32d)-(32e) S is a slack variable. The role of S is to set a hierarchy in the constraints and to ensure that the optimization problem returns a feasible solution [10]. For example, if the automated vehicle must be stopped to avoid the collision with a preceding vehicle and $v_{min,0} > 0$, the constraint (32e) cannot be guaranteed. It leads to the infeasibility of the optimization problem of (32). It must be avoided by setting

S to a high value. Since the avoidance of the collision has higher priority than keeping the speed is the predefined range, the following process must be performed for the selection of S. $S = 0$ is selected as a default value. If (32) has feasible solutions ρ_P, Δ_P, control input $u = \rho_P u_K + \Delta_P$ is computed. If (32) is not feasible with $S = 0$, S is selected for a high value to guarantee the feasibility. Then, (32) with the new value of S is solved. The resulted ρ_P, Δ_P are applied to provide control input $u = \rho_P u_K + \Delta_P$.

5 Design of the Robust LPV-based Cruise Control System

The aim of the robust LPV control is to provide u_K control input signal for the supervisor. The robust LPV control has importance in the situations, when the output of the predictive cruise control can violate the primary performances. Nevertheless, in most of the operation of the enhanced cruise control, u_P might be acceptable. Therefore, u_K has importance mainly in critical situations, and in these scenarios the maintenance of the secondary performances has low priority. Consequently, it is enough to use the simplified control-oriented model (18) for the robust LPV control design, with which the objective of the main functionality in cruise control, such as speed tracking can be specified as $z_1 = v_{ref,0} - \dot{\xi}, \quad |z_1| \rightarrow min$. Moreover, the minimization of the control input u_K must be considered as an objective of the robust control design: $z_2 = u_K, \quad |z_2| \rightarrow min$. The consideration of u_K has the role to guarantee the quantification of z_1 through the balance between the objectives. Furthermore, through z_2 the insufficiently high longitudinal control force is avoided. The objectives z_1, z_2 are composed in a vector of objectives, such as $z_K = \begin{bmatrix} z_1 & z_2 \end{bmatrix}^T$. Using the state-space formulation of the system (20), z_K is formed as $z_K = C_1 x + D_{11} w_K + D_{12} u_K$, where w_K is extended as $w_K = \begin{bmatrix} F_d & \Delta_P & v_{ref,0} \end{bmatrix}^T$, C_1, D_{11}, D_{12} are matrices.

The measurement equation for the robust LPV control design is formed as $y_K = v_{ref,0} - \dot{\xi} = C_2 x + D_{21} w_K$, where C_2, D_{21} are matrices. If $N^p = 0$, $v_{ref,0}$ is get from static map database of speed limits and the camera-based traffic sign recognition system of the vehicle. If $N^p > 0$, the speed information of database and the speed information of the recognition system are limited by radar measurement about $\dot{\eta}^p$. Thus, in the robust LPV control low number of external information is incorporated and most of the information is based on own sensors.

Finally, the plant for the robust LPV control design is formed as follows:

$$\dot{x} = Ax + B_1 w_K + B_2(\rho_P) u_K, \tag{33a}$$

$$z_K = C_1 x + D_{11} w_K + D_{12} u_K, \tag{33b}$$

$$y_K = C_2 x + D_{21} w_K, \tag{33c}$$

in which ρ_P is the scheduling variable of the system.

The control design is based on the resulted control-oriented model (33). Scaling of w_K and providing a balance between the elements of z_K require a weighting strategy in the control design method. The closed-loop interconnection structure is presented

in Figure (2). The interconnection structure contains several weighting functions. The

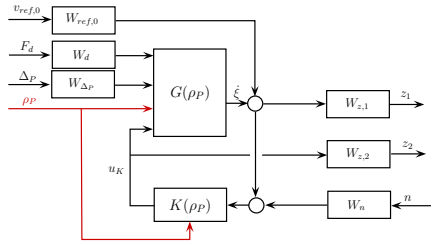


Figure 2
Closed-loop interconnection structure for robust LPV control design

weight W_n is related to the sensor characteristics on the velocity error measurement, where n represent sensor noise. W_d scales the longitudinal disturbance force F_d. The bound of F_d has also role in the supervisor design. Weight W_d is characterized as $W_d = \frac{F_{d,max}}{T_d s+1}$, where T_d is a tuning parameter, which represents the dynamics of F_d variation. Similarly, W_{Δ_P} scales the uncertainty Δ_P. This weight is selected in the form of $W_{\Delta_P} = \frac{\max(|\Delta_{P,min}|;|\Delta_{P,max}|)}{T_{\Delta 2}s^2+T_{\Delta 1}s+1}$, where $T_{\Delta 2}, T_{\Delta 1}$ are design parameters, which represent the dynamics of the signal. The role of weight $W_{ref,0}$ is to scale the reference signal $v_{ref,0}$. It is considered as a constant parameter with the supreme of $v_{max,0}$.

$W_{z,1}, W_{z,2}$ are the weights on the control performances, which provide a balance between them. Weight $W_{z,1}$ has important role from the aspect of the minimum performance level of the cruise control, because it scales the tracking error $v_{ref,0} - \dot{\xi}$. The form of the weight is $W_{z,1} = \frac{e_v}{T_z s+1}$, where T_z is a design parameter and e_v is the expected maximum tracking error. The selected form guarantees that the tracking error is e_v in steady state. The selection of e_v must guarantee that $e_v \leq v_{max,0} - v_{ref,0}$ and $e_v \leq v_{ref,0} - v_{min,0}$ to avoid the degradation of performance (2). Weight $W_{z,2}$ scales the control input u_K. Its value is selected as a constant parameter, which represents the supreme of $|u_K|$.

The quadratic robust LPV problem is to choose the parameter-varying controller $\mathscr{K}(\rho_P, y_K)$ in such a way that the resulting closed-loop system is quadratically stable and the induced $\mathscr{L}_2$ norm from the disturbance w_K to the objectives z_K is less than the value γ [11, 12]. The minimization task is the following:

$$\inf_{\mathscr{K}(\rho_P,y_K)} \sup_{\rho_P \in \rho_P} \sup_{\substack{\|w_K\|_2 \neq 0, \\ w_K \in \mathscr{L}_2}} \frac{\|z_K\|_2}{\|w_K\|_2}. \tag{34}$$

The existence of a controller that solves the quadratic robust LPV problem can be expressed as the feasibility of a set of LMIs, which can be solved numerically. Finally, the state-space representation of the robust LPV control $\mathscr{K}(\rho_P, y_K)$ is constructed [11, 13], which leads to the control input u_K. The input signal u_K is incorporated in the computation of u together with the selection of ρ_P, Δ_P. The control strategy results in that the minimum performance level of the closed-loop system is determined by

$\mathscr{K}(\rho_P, y_K)$. The computation of the robust LPV controller through the Matlab tool of [17] can be efficiently performed. Moreover, in the application-oriented papers [18, 19] further details on the computation and implementation of the robust LPV control can be found.

The optimization problem (34) shows that the resulting controller depends on the domains ρ_P, Λ_P, which demonstrates that the selection process of ρ_P, Λ_P and the robust LPV design are not independent from each other. During the control design it is necessary to find a balance in the selection of the domain, which is based on an iteration process.

The goal of the iteration is to find domains ρ_P, Λ_P, with which u approximates u_P (17). It provides that the enhanced cruise control system operates with u_P as most as possible, without the violation of primary performances. The following optimization is based on scenarios, which are performed in each steps of the iterations:

$$\min_{\substack{\rho_{P,min},\rho_{P,max}\\ \Delta_{P,min},\Delta_{P,max}}} \sum_{j=1}^{N} \big(u(j)-u_P(j)\big)^2 = \min_{\substack{\rho_{P,min},\rho_{P,max}\\ \Delta_{P,min},\Delta_{P,max}}} \sum_{j=1}^{N} \big(\rho_P(j)u_K(j)+\Delta_P(j)-u_P(j)\big)^2, \tag{35}$$

where j expresses the time step and N is the length of a given scenario.

The solution of the optimization problem (35) begins with domains with high ranges, which are reduced through the following iteration process.

1. The domain of the scheduling variable $\rho_P = [\rho_{P,min}; \rho_{P,max}]$ and the domain of the uncertainty $\Lambda_P = [\Delta_{P,min}; \Delta_{P,max}]$ are selected high in the first step, which can result in a conservative robust LPV controller.
2. The robust LPV control with the selected domains is designed using (34).
3. The closed-loop system with the incorporation of the designed $\mathscr{K}(\rho_P, y_K)$ and the domains ρ_P, Λ_P are analyzed through various scenarios. It yields in the signals λ and ξ, from which the cost in (35) for the scenario is calculated.
4. Due to the results of the scenarios the boundaries are modified to reduce the cost function of the optimization problem (35). The setting of the variables in the optimization can be performed through e.g., simplex search or trust region reflective methods, see [14, 15].
5. The robust LPV design, the scenarios and the evaluation (see steps 2-4) are performed until the cost (35) is higher than ε, where $\varepsilon > 0$ is a previously selected parameter.

The results of the entire iteration process are the robust LPV controller $\mathscr{K}(\rho_P, y_K)$ and the domains ρ_P, Λ_P.

6 Illustration of the Enhanced Cruise Control Strategy

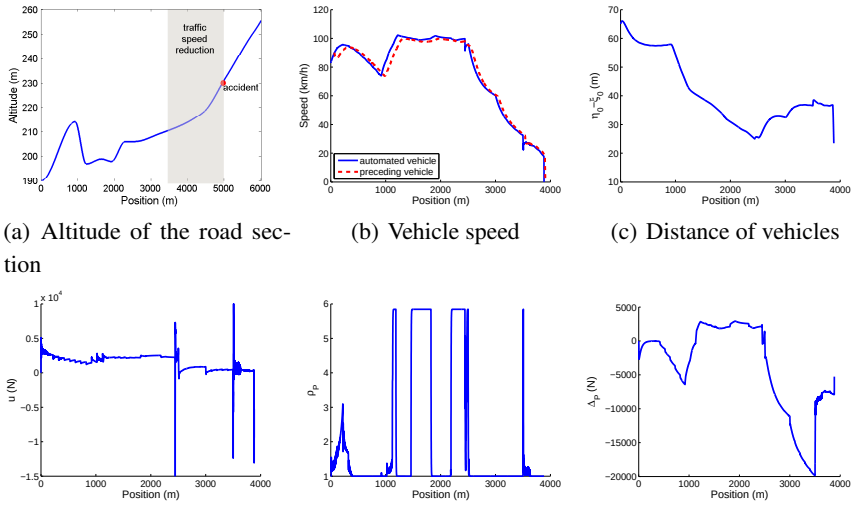
The effectiveness of the enhanced cruise control method is demonstrated in simulation examples. Two simulations are presented, which focus on the avoidance of primary

performance degradation in various scenarios.

6.1 Cruising on a Congested Highway

In the first simulation example the automated vehicle travels on a section of the hilly Hungarian M1 highway, which interconnects the capital cities Budapest and Vienna. In the example there is an accident on the highway (see $5000m$ segment point in Figure 3(a)), which results in a congestion. The traffic control system provides information about the reduction of the traffic speed between $3500m - 5000m$, see Figure 3(a).

The automated vehicle incorporates the traffic speed information in its enhanced cruise control strategy through $v_{ref,i}$ (see Section 2). Since the automated vehicle has $1000m$ prediction horizon in the example, the information about the reduced traffic speed is considered from $2500m$. It results in the reduction of the vehicle speed (see Figure 3(b)), with which an energy-efficient motion can be achieved through the approaching to the accident [7]. Moreover, there is a preceding vehicle ahead of the automated vehicle, which stops at $3800m$, when the congestion is reached, see its speed profile in Figure 3(b). The goal of the enhanced cruise control is to provide minimum control force in the cruising, while the safe distance $d_{safe} = 20m$ from the preceding vehicle is guaranteed, especially at the stop of the preceding vehicle. The distance between the vehicles is illustrated in Figure 3(c). It can be seen that the safety distance $20m$ is guaranteed in the given example. The control signal u and the values of ρ_P, Δ_P are illustrated in Figure 3(d)-(f). At the end of the simulation, when the distance is reduced, ρ_P is set to zero and Δ_P is also reduced. It results in the tracking of the preceding vehicle speed (close to zero) through the control input of the robust LPV control.



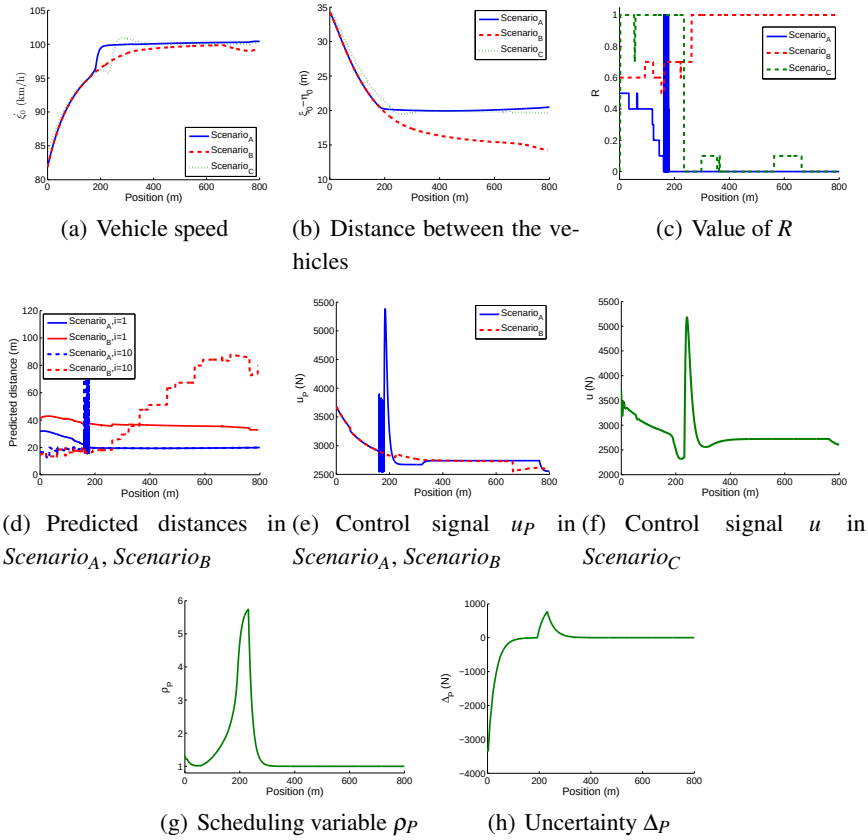
(a) Altitude of the road section (b) Vehicle speed (c) Distance of vehicles

(d) Control signal u (e) Scheduling variable ρ_P (f) Uncertainty Δ_P

Figure 3
Simulations of lane change scenario

6.2 Overtaking Scenario with Degradation in the Communication

The second simulation example presents a lane change scenario, in which the vehicle in the inner lane of a two-lane road overtakes a slower vehicle. There is also a preceding vehicle in the inner lane, which has higher speed, compared to the automated vehicle. In this situation the automated vehicle cannot change lane due to the overtaken vehicle, which is in the outer lane. Moreover, the vehicles in the inner lane also cannot be forced to reduce their speed, which means that the automated vehicle must be accelerated. The goal of the cruise control is to minimize the control force of the automated vehicle, while the safe distance between the vehicles is guaranteed.



(a) Vehicle speed (b) Distance between the vehicles (c) Value of R

(d) Predicted distances in $Scenario_A$, $Scenario_B$ (e) Control signal u_P in $Scenario_A$, $Scenario_B$ (f) Control signal u in $Scenario_C$

(g) Scheduling variable ρ_P (h) Uncertainty Δ_P

Figure 4
Simulations in overtaking scenario

In the example three scenarios are illustrated. In all scenarios the automated vehicle has information about the longitudinal acceleration, the speed and the position of the preceding vehicle through V2V communication, which is used in the predicted cruise control system to compute R, see (11). In $Scenario_A$ the time delay related to

the communication is $0.05s$, which is considered to be the nominal time delay value, while in $Scenario_B$ and in $Scenario_C$ the V2V communication has a degradation in the time delay, which is increased to $0.5s$. In $Scenario_A$ and in $Scenario_B$ the automated vehicle uses the presented predictive cruise control ($u \equiv u_P$), while in $Scenario_C$ the enhanced cruise control structure is in the loop, which also uses onboard distance measurement about the distance between the vehicles. The purpose of the simulation examples is to illustrate that the proposed enhanced cruise control method is able to guarantee the safe distance, even if the V2V communication delay is degraded.

Figure 4(a) shows vehicle speed and Figure 4(b) illustrates the distance between the vehicles in each scenarios. In $Scenario_A$ the predictive cruise control is able to guarantee the safe distance $20m$, which is resulted by the increase of its speed to $100km/h$. The increase of the speed is resulted by the reduction of R (Figure 4(c)), which is computed based on the reduction of the predicted distances, see e.g., section $0 \ldots 200m$ of $Scenario_A$ in Figure 4(d). It induces sharp increase in u_P, see Figure 4(e). Thus, the predicted cruise control is able to guarantee the safe distance, if the signals in the communication have low time delay value.

If time delay is increased, the predictions of the distances in $Scenario_B$ significantly differ from the predictions in $Scenario_A$, see Figure 4(d). Due to the increased time delay the preceding vehicle is predicted to have significantly smaller acceleration, which means that the automated vehicle focuses on the minimization of the control force. The increased R (Figure 4(c)) leads to reduced u_P (Figure 4(e)), which results in reduced speed profile (Figure 4(a)). Consequently, the safe distance between the vehicles is not kept (Figure 4(b)), which can force the preceding vehicle to unwanted braking intervention.

The resulted speed profile in $Scenario_C$ is illustrated in Figure 4(a). Since the supervisor uses the onboard measurement about the distance between the vehicles, the reduction in $\xi_0 - \eta_0$ is perceived. It leads to the reduction of ρ_P and Δ_P, with which the tracking of $v_{ref,0}$ is highlighted. The resulted control signal u of $Scenario_C$ (Figure 4(f)) is close to the u_P of $Scenario_A$, which results in keeping safe distance (Figure 4(b)). Thus, the enhanced cruise control system is able to guarantee safe distance, even if the time delay in the communication is significantly increased.

Conclusions

The proposed enhanced cruise control strategy is able to provide guarantees on the specified primary performances for automated vehicles. The consequence of the proposed method is that in most of the cruise control operation, requirements against the secondary performances (e.g., energy-efficient motion of the automated vehicle) can be maintained, while the primary performances are guaranteed in the entire operation of the control. The effectiveness of the control strategy is illustrated by simulation scenarios.

The proposed enhanced cruise control strategy is independent from the internal structure of the predictive optimal cruise control. Therefore, the future challenge of the method is its application to provide guarantees for further cruise control algorithms, e.g., learning-based cruise control methods.

Acknowledgement
The research was supported by the European Union within the framework of the National Laboratory for Autonomous Systems (RRF-2.3.1-21-2022-00002). The research was also supported by the National Research, Development and Innovation Office (NKFIH) under OTKA Grant Agreement No. K 135512.

References

[1] A. Sciarretta and A. Vahidi, *Energy-Efficient Driving of Road Vehicles*. Springer Verlag, 2019.

[2] J. Han, A. Vahidi, and A. Sciarretta, "Fundamentals of energy efficient driving for combustion engine and electric vehicles: An optimal control perspective," *Automatica*, vol. 103, pp. 558 – 572, 2019.

[3] E. Hellström, M. Ivarsson, J. Åslund, and L. Nielsen, "Look-ahead control for heavy trucks to minimize trip time and fuel consumption," *Control Eng. Practice*, vol. 17, no. 2, pp. 245–254, 2009.

[4] G. Wu, F. Ye, P. Hao, D. Esaid, K. Boriboonsomsin, and M. Barth, "Deep learning-based eco-driving system for battery electric vehicles," 2019.

[5] M. H. Almannaa, H. Chen, H. A. Rakha, A. Loulizi, and I. El-Shawarby, "Field implementation and testing of an automated eco-cooperative adaptive cruise control system in the vicinity of signalized intersections," *Transp. Research Part D: Transport and Environment*, vol. 67, pp. 244 – 262, 2019.

[6] P. Gáspár and B. Németh, *Predictive Cruise Control for Road Vehicles Using Road and Traffic Information*. Springer Verlag, 2019.

[7] B. Németh and P. Gáspár, "The relationship between the traffic flow and the look-ahead cruise control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1154–1164, May 2017.

[8] Z. Wang, Y. Bian, S. E. Shladover, G. Wu, S. E. Li, and M. J. Barth, "A survey on cooperative longitudinal motion control of multiple connected and automated vehicles," *IEEE Intelligent Transportation Systems Magazine*, vol. 12, no. 1, pp. 4–24, Spring 2020.

[9] A. Mihály, B. Németh, and P. Gáspár, "Look-ahead control of road vehicles for safety and economy purposes," *Control Conference (ECC), 2014 European*, pp. 714–719, 2014.

[10] S. M. Thornton and S. Pan and S. M. Erlien, and J. C. Gerdes, "Incorporating ethical considerations into automated vehicle control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1429–1439, June 2017.

[11] F. Wu, X. Yang, A. Packard, and G. Becker, "Induced L_2 norm controller for LPV systems with bounded parameter variation rates," *Journal of Robust and Nonlinear Control*, vol. 6, pp. 983–988, 1996.

[12] C. Briat, *Linear Parameter-Varying and Time-Delay Systems*, ser. Advances in Delays and Dynamics. Springer-Verlag Berlin, 2015.

[13] O. Sename, P. Gáspár, and J. Bokor, *Robust Control and Linear Parameter Varying Approaches*. Springer Verlag, Berlin, 2013.

[14] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112 – 147, 1998.

[15] T. Coleman and Y. Li, "An interior, trust region approach for nonlinear minimization subject to bounds," *SIAM J. on Optimization*, vol. 6, pp. 418–445, 1996.

[16] B. Németh and P. Gáspár, "Robust look-ahead cruise control design based on the $\mathscr{H}_\infty$ method," *IFAC-PapersOnLine*, vol. 48, no. 14, pp. 19 – 24, 2015, 8th IFAC Symposium on Robust Control Design ROCOND 2015.

[17] A. Hjartarson and P. Seiler and A. Packard, "LPVTools: A Toolbox for Modeling, Analysis, and Synthesis of Parameter Varying Control Systems," *IFAC-PapersOnLine*, vol. 48, no. 26, pp. 139-145, 2015.

[18] B. Németh, P. Gáspár, "LPV Design for the Control of Heterogeneous Traffic Flow with Autonomous Vehicles," *Acta Polytechnica Hungarica*, vol. 16, no. 7, pp. 233 – 246, 2019.

[19] G. Eigner, "Novel LPV-based control approach for nonlinear physiological systems," *Acta Polytechnica Hungarica*, vol. 14, no. 1, pp. 45–61, 2017.

Analyzing Narratives of Patient Experiences: A BERT Topic Modeling Approach

Mátyás Osváth[1], Zijian Győző Yang[2,3], Karolina Kósa[1]

[1]Department of Behavioral Sciences, Faculty of Medicine,
University of Debrecen
Nagyerdei krt. 98, H-4032 Debrecen, Hungary
osvath.matyas@med.unideb.hu; kosa.karolina@med.unideb.hu

[2]Hungarian Research Centre for Linguistics
Benczúr u. 33, H-1068 Budapest, Hungary
yang.zijian.gyozo@nytud.hu

[3]Sightspot Network Ltd.
Vágóhíd u. 2, H-4034 Debrecen, Hungary
yang.zijian.gyozo@nytud.hu

Abstract: Due to healthcare systems increased focus on healthcare quality and patient-centered care, the patients' perspective of delivered healthcare, has become an important part of healthcare service evaluations. Patient experiences can be used to improve the quality of care, as they reveal important information about health care encounters. An increasing number of organizations systematically collect and analyze patient experience data. The aim of our study was to identify major topics in narratives of patients' healthcare related experiences and analyze the reactions of readers of patient experiences. 1663 blogs and 298806 textual comments were extracted on non-solicited patient experiences from a Hungarian online forum during a 10-year period. Topic modeling with state-of-the-art BERT embeddings were used to analyze the data and extract meaningful patterns and concepts. Sentiment analysis was utilized to categorize the emotional valence of the narrative writings. The huBERT and HIL-SBERT models identified 326 and 200 topics in terms of patient experiences and 508 and 728 topics regarding the reactions to these experiences without human supervision. Conceptually similar topics were integrated into major categories with manual analysis. 94.4% of the experiences and 77.5% of comments were classified as negative, reflecting the same annual tendency over the decade. Our study uses a data-driven approach for extracting patterns of healthcare related patient opinions, in Hungary. Topic modeling, based on BERT embeddings, could provide useful information on patient perceptions and perspectives, that could improve healthcare quality and safety.

Keywords: NLP; topic modeling; sentiment analysis; patient experience; health care quality

1 Introduction

One of the indicators of health system performance is healthcare quality, which is high priority for the health policy sector and healthcare organizations aiming at patient-centered and quality care. Several theoretical frameworks – consisting of various indicators – have been developed for measuring healthcare quality. In 1966, Avedis Donabedian published a lasting framework and proposes to characterize the quality of healthcare with indicators that can be classified into three dimensions, namely structure, process and outcome [1]. The Organization for Economic Co-operation and Development (OECD) issued a framework as a nested-matrix aimed to measure the healthcare system performance, further developing Donabedian's quality of care [2]. One of the matrices essential dimension is patient experience, which includes every interaction between patient and the system healthcare, influencing the patients' opinion about healthcare. Amongst others, patient experience, comprises the communication between patient and healthcare personal, access to care and waiting time. It is important to distinguish "patient experience" from a different indicator called "patient satisfaction", which refers to patients' evaluation of their care compared to what they previously expected [3].

Along with the proliferation of the Internet, sharing experiences of and opinions on health services have "skyrocketed" in the form of reviews, blogs and comments. Patient online reports (POR) and patient report websites (PRW) allow users to review their or others health care encounters freely and anonymously in the form of quantitative ratings and/or textual feedbacks [5]. Analysis of non-structured texts and reviews can contribute to the improvement of healthcare quality and advance in person-centered measures [6]. Online reviews and comments could affect the attitudes and perceptions of other users and influence the decision making of patients [7] when selecting physician and healthcare services. According to a Hungarian study, at least 60% of the respondents chosen a physician based on online information and 17% of them regularly seek medical information online [8], which number is increasing.

Organizations increasingly collect and analyze online data anonymously with APIs and web scraping techniques. In the USA, several organizations operate web applications (RateMD, Healthgrades, Yelp) to collect and analyze patients' quantitative ratings and qualitative textual feedbacks describing the received care in a certain hospital and department. The British National Health Choices (NHS) maintains an online portal for writing narrative feedbacks [9]. In Hungary, studies mostly used such survey methods to assess patients' experiences [4]. Studies suggests an association between positive patient experiences with less frequent use of secondary and tertiary care (e.g., hospital admission and readmission, consultations) [10]. Improved adherence, lower mortality, lower readmission rates and shorter inpatient stays characterizes hospitals with higher patient satisfaction scores [6].

Although the application of quantitative methods is dominant in measuring patient satisfaction and experiences, these methods are necessarily limited due to information loss and less intelligible and practical to healthcare personal. Gallan et al. (2016) concluded discrepancy between ratings and comments, i.e., highlighting the possible contrast between survey scores and textual feedbacks [11]. Text analysis allow a more detailed and fine-grained processing but requires more complicated techniques. Manual content analysis and annotation has been the standard way to analyze narratives in any domain [9] [12] [13], but practically impossible to extract information from large quantities of text. In the past decades, progression in Natural Language Processing (NLP) made the analysis of large volumes of text not only possible but scalable and proved to be an alternative to be used for information extraction and identifying valuable patterns from non-structured texts.

1.1 Related Work

Applications of Machine Learning (ML) and Natural Language Processing (NLP) to data related patient experiences face several challenges, for instance the volume, type of corpus, classification models used and overfitting [14]. Greaves et al. [15] utilized sentiment analysis and machine learning to classify patients' comments, and ML algorithms to predict different categories, such as hospital cleanliness, whether patients were treated with respect, and hospital recommendation. Doing-Harris et al. [16] extracted latent topics in patient comments about health care services using a machine learning based approach. Among others, their most frequent categories were related to appointment access, empathy, and explanation. Li et al. [5] developed a taxonomy for physician reviews including 3 domains related to physicians, patients and systems as well as 9 subcategories. A mixed-method approach was used including literature review, human annotation, and topic modeling.

In 2021, a systematic review summarized studies that applied ML and NLP models to patient experience to identify patterns from non-structured texts [12]. Studies were categorized based on supervised, unsupervised, semi-supervised, rule-based or dictionary-based approaches. As the study highlighted, a common method utilized is sentiment analysis, which is used to detect emotional valence expressed by the writer in the text, classifying texts into negative, positive or neutral category. All studies with unsupervised approaches (n=6) used topic modeling, i.e., to discover latent semantic structure and automatically identify topics in a collection of documents and texts. Different algorithms can be used to recognize patterns and give new insights within data. However, most studies relied on Latent Dirichlet Allocation (LDA) algorithm for topic modeling, which – according to Angelov [17] – has several disadvantages including the need to define the number of topics a priori and disregarding word semantics. We find one study, where contextual information from clinical texts of patients – such as disease severity, time and onset

– were extracted by using state-of-the-art BERT embeddings. Their experimental result achieved higher accuracy rate than with Word2vec model [18].

According to our knowledge, this is the first study in Hungary that focuses on analyzing narrative patients' experiences as well as reactions to these experiences by other readers.

1.2 Objective

We hypothesized that non-structured blogs and users' responses to them in the form of comments provide meaningful information on patient experiences, which can be extracted by NLP techniques, getting insight from patients' and users' texts. The aim of the study was to identify patterns and topics in non-solicited patient experiences about Hungarian healthcare system posted to an online forum. Furthermore, readers' comments on the spontaneous narrative experiences were examined to understand the interactions in the comment section. To encode the text of the comments, two state-of-the-art contextualized BERT models were applied, and unsupervised machine learning algorithm were utilized to cluster the results. Additionally, sentiment analysis (with a fine-tuned BERT model) was used to categorize the valence (positive, negative, and neutral) of non-structured texts.

2 Methods

2.1 Data Source

Data were extracted from an online forum (https://praxis.blog.hu/) dedicated to sharing patient experiences in Hungarian health care. The non-institutional website was established in 2009, and the daily number of visitors ranges from 5000 to 7000. Spontaneous descriptions of health care encounters emailed to the staff by patients, or their relatives are posted without editing anonymously in the form of blogs. Contents on the website are organized by week and year. Anyone can comment under each post and answer to a comment. Blogs and comments do not have word limits. Blogs are uploaded voluntarily and anonymously so that the identity of the writer or persons depicted in the blog cannot be established. Approval and informed consent were not necessary as the website from which blogs were downloaded is accessible to anyone and can be freely used for non-commercial purposes by referencing the source according to the Creative Commons License of Hungary (CC BY-NC-ND 2.5 HU) [38].

Our research focuses on the blogs and comments, with a total of 1663 blogs and 298.806 comments on patient experiences posted between 2009 and 2020. Data

were extracted by a Web crawler, and we deleted any information on personal identity or sensitive data. The final cleaned dataset comprises 1660 blogs and 267631 comments.

2.2 Data Analysis

Word embeddings were used to convert comments into numerical data. These are vector representation of words in a text, capturing contextual hierarchy. From that, similarity metrics can be calculated using distance measures (e.g., cosine distance). Deep learning models, such as Word2Vec are often used for extracting features and capturing context from corpus, but problems may arise in a highly specific context. Bidirectional Encoder Representations from Transformer (BERT) [19] resolves this issue by learning the contextual features between words, embedding into a continuous vector space, reading input text sequentially from right-to-left and left-to-right. BERT models can be pre-trained on large corpora of text and fine-tuned to be used in different contexts.

Our study builds on the prior work of Angelov [17] which applies state-of-the-art Transformer based word embeddings for topic modeling. We used two contextualized BERT embeddings to encode the text of the comments. For the first BERT model, namely huBERT, preprocessing of the input text were not necessary, except for tokenization, as the model was trained similarly on a 9-billion-token Hungarian Webcorpus 2.0 [20], outperforming other BERT models in various NLP tasks [21]. The architecture of huBERT is the same as BERT base model with 12 Transformer layers with 12 heads and 768 hidden dimensions [22]. Additionally, a fine-tuned HIL-SBERT model [23] was also applied for feature extraction which is based on Sentence-BERT [24], a modification of BERT model aiming to achieve high-quality sentence embeddings.

To improve the results of unsupervised clustering algorithm used for topic identification, dimension of the data was reduced using the Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) algorithm [25]. Next, the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm [26] was applied to the encoded low dimensional vector representation to capture the structure of the data automatically with the aim of grouping words to form topics. HDBSCAN is a density-based hierarchical clustering algorithm that does not require pre-defined number of clusters providing optimal solution for the reduction of problems with vector sparsity and computation costs. Models were fine-tuned to have 15 words per topic. Clustered results were lemmatized, and stop words and punctuation marks were removed for easier interpretation. Topics with null weights and duplicated topics were removed from the output. For interpreting the clustered results, c-TF-IDF (class-based TF-IDF) [27] were used for automatic topic identification. The final output of the pipelines were clusters, consisting of 15 words per cluster, and human interpretation was used

to assign labels for the topics and merged some of the topics to form more interpretable categories.

Sentiment analysis was utilized to identify the emotional content of the comments. To fine-tune the pre-trained huBERT model, the Hungarian Twitter dataset [28] – created by PreCognox [29] – was used. Two subcorpora were created as defined in the international benchmark [30] and applied to predict the class label of comments:

2-class: Class label 1 and 2 were converted to negative, class 4 and 5 to positive, ignoring 3 for ambiguities. Training corpus: 2468 segments. Test corpus: 269 segments.

3-class: Class label 1 and 2 were converted to negative, class 4 and 5 to positive and class 3 to neutral. Training corpus 3600 segments. Test corpus: 400 segments.

5-class: Classes were labelled as in the original Likert scale with 1 (very negative) to 5 (very positive).

Fine-tuning was performed with the following hyperparameters: batch size 8 / GPU (4 GeForce GTX 1080Ti); learning rate: 2e-5; sequence length of 128 and 15 epochs (models with best performance were used). The model achieved an accuracy of 85.92% for the binary, 72.18% for the 3-class corpus and 68.50% for the five-class variant.

3 Results

3.1 Topics Modeling using Patient Experiences

Analysis revealed that the resulting clusters with fine-tuned BERT models form semantically adequate topics. The encoded patient experiences analyzed by the huBERT and HIL-SBERT models resulted in 326 and 200 topics, respectively. Both models only use part of the corpus, 20503 sentences by the huBERT, and 20502 by HIL-SBERT model, clustering the rest as noise (huBERT: 33744, HIL-SBERT: 34745). Dimensionality reduction was applied to visualize the clustered results and show the representation of comments in continuous vector space.

No quantitative metrics were found to evaluate the unsupervised models; therefore, results were interpreted manually. The outcome of huBERT model had easier interpretation, i.e., resulted in more interpretable and semantically meaningful categories, while HIL-SBERT model had more noise in the results. Therefore, during the rest of the analysis we used to topics identified by the huBERT model. Irrelevant topics were discarded from the 326 topics. Human interpretations were applied to merge topics into higher level categories. To reduce bias, authors

discussed topics iteratively. Sentences in patient experiences are highly dependent to the specific situation in healthcare and to additional topics in the patient experience, as a narrative could consist of many topics. Considering their large number, only the five most notable categories will be described below, along with 18 topics marked important by the model (see Tables 1 and 2).

Diagnosis, Symptoms and Illnesses

This category comprises patients' or their relatives' perceived symptoms, illnesses along with their diagnosis. Symptoms were related to acute and chronic diseases ranging from influenza to cancer. Multiple topics were included in this category, as the model recognized multiple classification of diseases. This category captured dental, spine and ophthalmic problems, sexually transmitted diseases and plastic surgery. Expressions related to pain may indicate the difficulty of their condition and coping with pain. One patient wrote that: "…, pain was unbearable, as they did not anesthetize me…", another patient mentioned: "The diagnosis was a suspicion of purulent meningitis, probably correctly…"

Birth and Gynecology

Experiences related to gynecology, childbirths and birth were well differentiated by the BERT model. Recurring occurrences of topics has grounded the grouping of topics into one category. Category included the following topics: gynecological examinations, childbirth, pregnancy, abortion, infant, new-born, vaccination and protection. The model also identified topic related to postpartum depression. Expressions in this category included sentences like "Paradoxically, postpartum period doesn't bring happiness into mothers' lives." and "The child-birth itself went smoothly, the doctor was fair."

Family and Children

This category referred to the topics, which were likely to appear in the lives of families including hospitalization of the child, caring for the child, family and parenting roles, vaccination, and the condition of other relatives. It should be highlighted that several blog posts also reported events at the school and with the school's physician. The reason for separating this category from the former is that the former focuses on the mother, while words related to other relatives and children appear more frequently in the latter topics. One relative expressed: "A child must be treated with patience and attention, especially if he is ill."

Structure

Patients discussed multiple topics which can be classified into structure category forming the basis of the healthcare system and is consonant with Donabedian's dimension of structure. Topics identified by the model include problems with professionality, lack of human resources, relationship between healthcare and politics, and the malfunctioning of healthcare system. Multiple references were made regarding the inadequate hygiene circumstances in the hospitals and

departments. One patient articulated as follows: “I was thinking of sharing this with you because now there is chaos, bankruptcy, lack of physicians everywhere.”

Outcomes

Both favorable and unfavorable expression related to outcomes were included in this category. Sentences in unfavorable outcomes comprise dying, death and mourning and the uncertainty of recovery. The model identified several inauspicious outcomes of healthcare. One relative wrote the following: "The family believes there was an omission in the hospital and the overdose of medication led to the death of our loved one." Favorable outcomes included expression related to topics like healing and gratitude, or references to quality care provided by the healthcare personnel. One patient expressed: “A thousand thanks to those who are human despite all the difficulties.”

Table 1
Meaningful categories of patient experiences

#	**Category**	**Keywords**
1	Diagnosis, symptoms and diseases	spine, waist, toothache, herpes, pain, disc herniation, infusion, complaint, sclerosis, numbness, double vision, muscle pain
2	Structure	lack of nurses, disorganization, lack of doctors, remorse, care system, unfair, professional, heedlessness, high degree,
3	Outcome	work, class, grateful, love, nurse, working, bond, heart, mourning, man, mourning, death, dead, process
4	Birth and gynecology	mom, gynecologist, crib, childbirth, childbirth, abortion, slut, vaccination, protection
5	Family and children	toddler, bed, intelligence, need, development, child, father, role, mother, tongue, parent, carrier

Table 2
Keywords and their classification

#	**Topic**	**Keywords**
1	Medical Oath	Oath, Hippocratic, if, medical, worthy, insight, part, recognition, fulfil
2	Medicine and Drugs	Frontin, hormone, escitil, ingredient, acetyl, sore throat, ibuprof, tablet, paracetamol, agent, tablet
3	Vaccination	Vaccination, vaccination, against, influenza, protection, vaccination, pharmacy, administration
4	Blood	Blood, blood donation, blood donation, call, potential, blood pressure
5	Alternative medicine	Magneter, long, effect, initial, alternative, deterioration, cure, various, unexpected, improvement
6	God and science	God, eso, science, extreme, medicine, esoteric, pedestal,
7	Plastic intervention	Plastic, horrible, surgery, chest, detail, week, order, tube, cm, pair, surgeon, breast augmentation

8	Other healthcare services	Soothing, dispatcher, ambulance, smile, patience, sick, car, soothe, nurse, professional
9	Holidays and religion	Wreath, advent, candle, stove, Christian, mood, mythology, decorate, pine tree
10	Healthy lifestyle	Diet, food, dietary, varied, seven, nutritional, magnesium, stress, workplace, body, magnesium deficiency, b6, stressful
11	Workout	Workout, movement, amplifitness, proper, training plan, fitness room, muscle group, body awakening, device
12	Side effect(s)	Side effects, truxal, generalized, rulid, psychotic, paranoid, anxious
13	Smoking	Smoking, act, experience, anyone, doctor, many, get used to, addiction
14	Communication about health condition	Doctor, condition, communicate, ct, brother, how many, day
15	Referral	Date, examination, request, sheet, referral, give, know, doctor
16	Sharing experience	Story, love, share, experience, tell, blog, hospital, write, describe
17	Money	Tb, pay, ft, tax, disability benefit, aid, salary, support, money
18	Waiting times and appointment	Three, seven, days, after, hour, date, get, year, surgery, test, also, takes, waiting, line, after, half, year, waiting

Additional topics are shown in Table 2. Patients and their relatives discussed many topics in relation to health and health care such as blood, vaccination, medicine and their side effects, alternative cures, religion and science, smoking, lifestyle and workout. Further research is needed to answer the various questions arisen because of the uniqueness of these topics such as the roles of alternative cures, lifestyle, and the connection regarding the role of religion in science.

3.2 Topics Related to Users' Reaction (comments)

In contrast to the blog post topics based on the experiences of patients or relatives, users' reaction to experiences of patients mainly relies on the context. Linguistic characteristics reflected in comments contains meaningful information for determining the topic of feedback and the overall reactions to blog posts. Similarly, the same topic modeling pipeline were applied for the comments with c-TF-IDF for automatic topic identification and human interpretation for further processing.

The topics of comments resulted in four main categories, including symptoms, ethical issues, addressing feelings and consolation (Table 3). Symptoms, diseases and diagnosis category were mainly similar to the aforementioned category with the same name, referring to comments acknowledging the significance of specific symptoms of bloggers and commenters. The ethical issues category consists of

several subjects of ethical relevance including police, accusation and lawyer reveals a growing interest in the consequences of inappropriate healthcare, reflecting the available actions after an unfavorable outcome. The category of addressing feelings captures commenters feelings and emotions to other patients' experiences in healthcare. As different scenarios may trigger different feelings in the commenters, they may reply with anger to the causative factor or person and feel empathy toward the patient receiving inappropriate service. Expressions in consolation category mainly containing understanding and get-well messages signaling consolation and condolences. Users may show empathy and sympathy towards a patient whose encounter could end in unsatisfactory outcome.

Table 3
Meaningful categories of comments to patient experiences

#	**Category**	**Frequent words**
1	Symptoms	vision, mental disorders, chest pain, insulin, pain, virus, bacteria, cancer, bone, knee injury
2	Ethical issues	imprisonment, police, accusation, lawyer, theft, rules, protocol, patient representative, informed consent, custodial, physical, years, public interest, insult, danger, crime, violence, negligence, report, call, legal, summoned, witness evidence, criticize on personal grounds, house rule, state, justice, required, law
3	Addressing feelings	love, grief, anger, sadness, depression, rough, argument, outrageous, gas, scandal, unpleasant, fuhh, bloodthirsty, annoying, offended, back, outrageous attitude, fucking good, terrible, bad, disgusting, rude
5	Consolation	Sympathy, get well, condolences, apologies, my condolences, accept, feel, together, done happened, dear, get well, calm, rest in peace, settle down, meditation, error

3.3 Sentiment Analysis

Patient Experiences

Categorization of patient experiences by the fine-tuned huBERT model for the 10-year period revealed that the valence of sentiments was dominantly negative (94.4%) in the five-class variant. Similar patterns were showed in the other two variants. In the five-class model, the score 3 (neutral score) were used as cut-off point for deciding the polarity of comment (positive, negative). One patient experience could contain multiple emotional expressions, whereas different aspects and processes of care may lead to different patient perceptions, and it is possible that the same hospital care may result in a positive experience for one patient and a negative experience for another. The same tendency showed when segmented the analysis of the experiences by each year in the 10-year period, that is, most of the patient experiences were negative (Fig. 3).

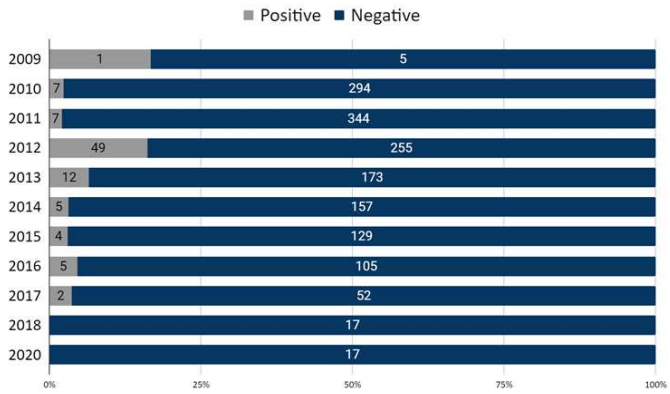


Figure 3
Sentiment analysis results during the 10-year period for patient experiences

Trend analysis were used to capture the changes in the emotional patterns in patients' experiences during the 10-year period. The sentiment score in the classified sentences were averaged and visualized sorted by year. As shown in Fig. 4, the majority of blogs reflect negative sentiment. Similar pattern can be seen in Fig. 5, which shows the distribution of the summarized sentiments of the classified sentences plotted on a five-point Likert scale.

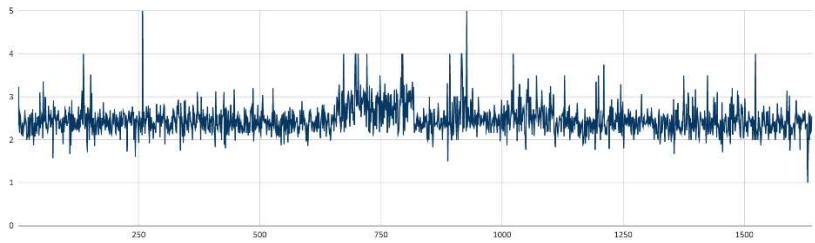


Figure 4
Distribution of blog posts by an average sentiment score sorted by years

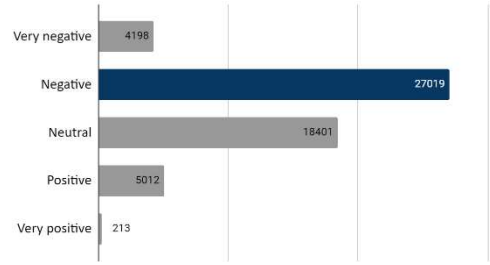


Figure 5
Overall distribution of the (blog posts') classified sentences

Comments

Similarly, the results of topic modeling showed the same tendency with the fine-tuned huBERT model for the 10-year period in the categorization of users' reactions to experiences, revealing dominantly negative (77.52%) sentiment in the binary (2-class) variant. Analysis by each year in the 10-year period showed the similar pattern, that is, the majority of sentiments in the comments were negative (Fig. 6). Commenters expressed mixed emotions (e.g., love, anger). Users may express their feelings and emotion towards the putative factors or persons that could have caused negative patient experience, not necessarily towards the specific patient, which may explain our results.

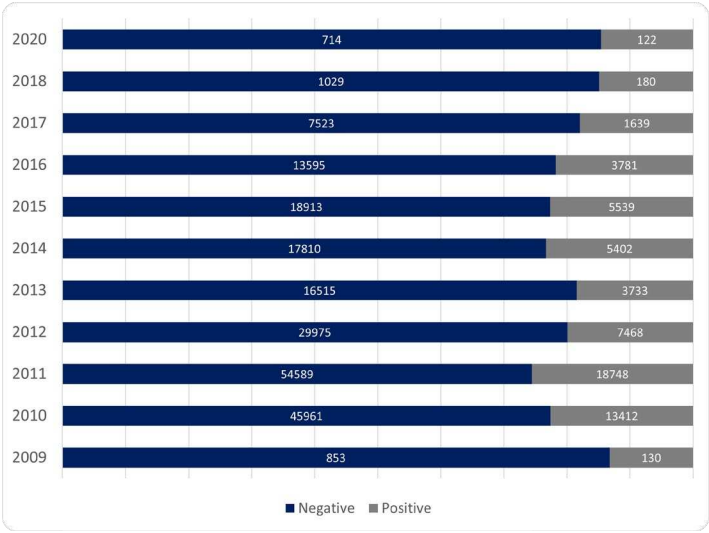


Figure 6
Sentiment analysis results segmented by year for the comments

4 Discussion

Our exploratory results present the first NLP-assisted analysis of narratives of patient experiences and reaction of various persons to narratives, emphasizing and reflecting mostly negative experiences occurring in Hungarian health care. Understanding the attitudes and opinions of patients is crucial as experiences reflects healthcare quality and may affect others' opinions and decision making related to the use of health care services. Furthermore, sharing opinions on the internet are growing and patients' views about health care services are becoming like other paying services.

Categories and topics identified in the narratives shows the frequently occurring problems healthcare considered important by patients. Our results revealed that five categories emerged from the analysis, namely symptoms and illnesses, structure, outcome of care, family and children, and birth and gynecology. Two categories are related to the classical theoretical framework describing quality of healthcare by Avedis Donabedian [1]. Other research has shown that satisfaction is correlated with system and management issues and coordination of care, for instance waiting time, access to care, hygiene and environment [9], which all contribute to negative experience if the service is unsatisfactory. The category of outcome of care were also relevant in our analysis. It is important to mention the subjective perception of care may differ between patients, as one patient could experience the treatment as suboptimal, while the other may rate the same treatment satisfactory. Additionally, other factors – such as previous illnesses, lifestyle factors, adherence – could have an impact on the outcome of care and were, "unfortunately", unknown in the blog posts.

The category of symptoms, diseases and diagnosis were prominent in both blog posts and comments, referring to the acknowledgement and significance of specific symptoms and diseases. Identifying and classifying various symptoms and diseases in narratives based on neural networks could contribute to the improvement of Clinical Named Entity Recognition systems. Moreover, a recent study proposed a framework to understand the contextual information of the parent entity (e.g., headache) to predict other symptoms' characteristics like time and severity, achieving superior performance against other state-of-the-art models [18].

Comments are mainly related to the specific context of health care encounter to which they are related. Our results show that – besides symptoms and disease category – three other categories emerged from the users' reaction to patient experiences. First, addressing feelings category contains mixed emotions. According to Table 3, there were many words related love, grief, anger and sadness. The accumulation of emotional words could help analyze the provoking factors behind negative experiences and develop a more personalized system focusing on negative patient feedbacks. Second, the category of ethical aspects and the violation of patient's right are one of the most essential factors to decrease satisfaction. The category consists of comments such as "The hospital must be sued…" or "You have to go back with a lawyer and ask for every documentation possible!".

Several topics identified by topic modeling help understand the various aspects and nuances of patient experiences and comments. Waiting times for appointment and care, delayed operations and difficulty of booking an appointment all contributes to negative experiences. According to Fernandes et al. [31] waiting times for in healthcare services differs in Hungary, the majority of respondents were seen within a week by family doctor, public specialist, or a private specialist; and only one third of patients waited more than one month. Based on our results, medicine and side-effects were important topic to patients, as many patients could have side effects after treatment [32]. Moreover, monitoring patients' feedbacks about potential side-

effects, and what to do in case of specific problems develop could increase patients' satisfaction. Establishing a relation between drugs and side-effect could be included in future research.

Our results revealed that the majority of experiences (94.4%) and comments were negative (77.5%). The same tendency shows when texts were segmented by year (see Fig. 3 and 6), contrary to other studies findings [9] [13] [33]. By taking advantage of previously created Hungarian Twitter dataset, we have been able to fine-tune the BERT model and use more learning set to achieve higher level of accuracy. Commenters expressed similar feelings as in the patients' writings, reflecting the negative experiences in health care. Nonetheless, the same user can react both positively and negatively to a specific patient experience, given that it could comprise multiple healthcare encounters, it makes more difficult for the sentiment algorithm to provide accurate results.

To our knowledge, one previous Hungarian study assessed patient experiences on a national representative sample (n=1000), which included several aspects of care highlighted by the OECD framework including patient-doctor communication, shared decision-making, time spent with patients and giving opportunity to ask questions. According to their research, all categories were dominantly positive (>80%), but lower when compared to other OECD countries [4] [31].

Our analysis showed a much higher proportion of problems with healthcare services (e.g., structure and waiting times). Although our source platform is skewed towards unsatisfied and frustrated patient experiences and they are do not accurately present the quality of care in Hungary, it proved to be a useful information source as it emphasizes the categories and topics which patients finds useful.

A psychological phenomenon, namely the negative bias could further explain why patients with unsatisfied experience are more likely to write about their frustration. Negative experiences are perceived to be more complex than positive ones and therefore result in more complex cognitive representations compared to positive ones. Since negative stimuli carry greater informational and adaptive value than positive stimuli, patients spend more time dealing with negative rather than with positive stimuli [34]. The majority of negative sentiments may refer to that those patient post on this online forum that could not or do not want to submit a complaint to patients' representatives in the legal way that exists in healthcare. Nevertheless, negative experiences provide greater informational value compared to positively valenced or neutral experiences because inferences can be drawn for the improvement of health care quality.

As to the limitations, it is important to mention that our source of data is not representative of the Hungarian healthcare, therefore generalizability is limited for drawing conclusions. Unsolicited online reviews have a higher probability of selection bias as unsatisfied patients are more likely to complain. Moreover, accounting for the more than one million nursing days in all inpatient facilities in Hungary, this data with mostly negative patient experiences submitted in 10-year

period could be considered small [35]. Another limitation is human interpretation used to evaluate models' performance as no quantitative metrics were found for the unsupervised models. Differentiation between the identified topics in the blog posts poses challenges, as a patient can mention different problems in one blog, and one problem could spread across sentences and paragraphs. Further research is needed to clarify methodological details such as the optional number of topics and differentiation between topics to provide more accurate results. Additionally, sentiment analyses were only applied to temporal categories, i.e., to blogs and comments divided by years, but not to the topics themselves. Further analysis is needed to assess the polarity of the identified topics and categories.

The Hungarian healthcare system provides health coverage for almost all its citizens based on universal health insurance. However, it is less exhaustive than in other EU countries [55], and 22.5% of the population reported unmet medical needs due to various difficulties (e.g., financial, waiting list, distance/transportation) in 2014 which is below the EU average. These factors can affect patient satisfaction and decision making, which could increase the possibility for complaints. Rules for investigating patient complaints are specified in the international regulation of healthcare providers that are also covered by national regulations. Patients can log complaints with the representatives in person, by phone, mail or email. According to the official website, studies evaluating patient satisfaction showed approximate satisfaction rates around 90% with 10% of negative feedback [36], in contrast to our findings. However, lack of information as to where and how to file an official complaint and/or lack of trust in official platforms prevented some patients from complaining formally. Several patients reported supposed uninterest from the representative of patient rights feeling that "they will not be taken seriously". While both providers and patients may fear consequences and detrimental effects [37], reporting complaints and responding to them online would help in a more direct, open communication, allowing providers to be better informed and implement necessary steps to improve their services [5].

The value of our research is that our findings potentially reveal the patterns and topics in patient experiences and directs attention on the examination of spontaneous non-solicited narratives, providing an opportunity to explore situations and encounters that probably cause the most frustration and dissatisfaction among patients. Most of the topics are related to limitations, shortcomings and deficits in health care services. The analysis of patient experiences – specially in large volume – could provide an important addition in improving the quality of care, especially if the analysis could be extended to names and localizations.

Our exploratory results showed the usefulness of BERT topic modeling to automatically process patient narratives. It highlights the need to develop an official patient online review web application embedded in Hungarian healthcare system to provide effective and continuous monitoring, which could contribute to the improvement of healthcare quality.

Conclusions

Our research aim was to automatically analyze and extract patterns of patient experiences and comments, made on the experiences of patients over a decade. Topic modeling, with BERT embeddings, were applied to identify frequently occurring topics and categories from both blog posts and comments. Five major categories were related to narratives, two of them – namely, structure and outcome category – were in consonant with previous theoretical frameworks for quality of care, and four categories emerge from the comments, respectively. Analysis was augmented by human interpretation to evaluate unsupervised models' performance, and to merge similar topics into categories. Sentiment analysis were used to categorized texts by polarity (positive, negative and neutral). The majority of texts were classified negatively by the model, implying dominantly negative health care encounters. To understand physician-patient encounter and problematic topics of healthcare services, NLP methods and techniques can be utilized effectively. However, the methods applied in our research, must be refined and improved, to strengthen the validity of our results and achieve less manual interpretation.

Funding

This research was funded by EFOP-3.6.1-16-2016-00022 "Debrecen Venture Catapult programme" project (co-financed by the European Union and the European Social Fund).

References

[1] A. Donabedian, "Evaluating the quality of medical care," Milbank Quarterly, Vol. 83, No. 4, Blackwell Publishing Inc., pp. 691-729, 2005, doi: 10.1111/j.1468-0009.2005.00397.x

[2] F. Carinci et al., "Towards actionable international comparisons of health system performance: Expert revision of the OECD framework and quality indicators," Int. J. Qual. Heal. Care, Vol. 27, No. 2, pp. 137-146, Apr. 2015, doi: 10.1093/intqhc/mzv004

[3] E. Larson, J. Sharma, M. A. Bohren, and Ö. Tunçalp, "When the patient is the expert: Measuring patient experience and satisfaction with care," Bull. World Health Organ., Vol. 97, No. 8, pp. 563-569, Aug. 2019, doi: 10.2471/BLT.18.225201

[4] Ó. Brito Fernandes et al., "Patient experiences with outpatient care in Hungary: results of an online population survey," Eur. J. Heal. Econ., Vol. 20, No. Suppl 1, pp. 79-90, Jun. 2019, doi: 10.1007/s10198-019-01064-z

[5] J. Li, M. Liu, X. Li, X. Liu, and J. Liu, "Developing embedded taxonomy and mining patients' interests from web-based physician reviews: Mixed-methods approach," J. Med. Internet Res., Vol. 20, No. 8, Aug. 2018, doi: 10.2196/jmir.8868

[6] K. Nawab, G. Ramsey, and R. Schreiber, "Natural Language Processing to Extract Meaningful Information from Patient Experience Feedback," Appl. Clin. Inform., Vol. 11, No. 2, pp. 242-252, 2020, doi: 10.1055/s-0040-1708049

[7] C. M. Burkle and M. T. Keegan, "Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians," BMC Health Serv. Res., Vol. 15, No. 1, pp. 1-7, Sep. 2015, doi: 10.1186/s12913-015-1099-2

[8] V. Zsuzsa and H. Tamás, "Patients' preferences for health-related use of Internet" Orv. Hetil., Vol. 159, No. 51, pp. 2175-2182, 2018, doi: 10.1556/650.2018.31210

[9] T. Lagu, S. L. Goff, N. S. Hannon, A. Shatz, and P. K. Lindenauer, "A mixed-methods analysis of patient reviews of hospital care in England: Implications for public reporting of health care quality data in the United States," Joint Commission Journal on Quality and Patient Safety, Vol. 39, No. 1, Joint Commission Resources, Inc., pp. 7-15, Jan. 01, 2013, doi: 10.1016/s1553-7250(13)39003-5

[10] R. A. Price et al., "Examining the role of patient experience surveys in measuring health care quality," Medical Care Research and Review, Vol. 71, No. 5, SAGE Publications Inc., pp. 522-554, Oct. 08, 2014, doi: 10.1177/1077558714541480

[11] A. S. Gallan, M. Girju, and R. Girju, "Perfect ratings with negative comments: Learning from contradictory patient survey responses," Patient Exp. J., Vol. 4, No. 3, pp. 15-28, 2017, doi: 10.35680/2372-0247.1234

[12] M. Khanbhai, P. Anyadi, J. Symons, K. Flott, A. Darzi, and E. Mayer, "Applying natural language processing and machine learning techniques to patient experience feedback: A systematic review," BMJ Health and Care Informatics, Vol. 28, No. 1, BMJ Publishing Group, p. 100262, Mar. 02, 2021, doi: 10.1136/bmjhci-2020-100262

[13] A. López, A. Detz, N. Ratanawongsa, and U. Sarkar, "What patients say about their doctors online: A qualitative content analysis," J. Gen. Intern. Med., Vol. 27, No. 6, pp. 685-692, 2012, doi: 10.1007/s11606-011-1958-4

[14] A. Le Glaz et al., "Machine Learning and Natural Language Processing in Mental Health: Systematic Review," J. Med. Internet Res., Vol. 23, No. 5, May 2021, doi: 10.2196/15708

[15] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, and L. Donaldson, "Use of sentiment analysis for capturing patient experience from free-text comments posted online," J. Med. Internet Res., Vol. 15, No. 11, pp. 1-9, 2013, doi: 10.2196/jmir.2721

[16] K. Doing-Harris, D. L. Mowery, C. Daniels, W. W. Chapman, and M. Conway, "Understanding patient satisfaction with received healthcare

services: A natural language processing approach," AMIA ... Annu. Symp. proceedings. AMIA Symp., Vol. 2016, pp. 524-533, Jan. 2016, Accessed: Nov. 17, 2020 [Online] Available: https://europepmc.org/articles/PMC5333198

[17] D. Angelov, "Top2Vec: Distributed Representations of Topics," arXiv, Aug. 2020, Accessed: Apr. 27, 2021 [Online] Available: http://arxiv.org/abs/2008.09470

[18] B. Saha, S. Lisboa, and S. Ghosh, "Understanding patient complaint characteristics using contextual clinical BERT embeddings," 2020

[19] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf., Vol. 1, No. Mlm, pp. 4171-4186, 2019

[20] D. M. Nemeskey, "Natural Language Processing Methods for Language Modeling", Doctoral Dissertation, Doctoral School of Informatics, Faculty of Informatics, Eötvös Lóránd University, 2020

[21] D. M. Nemeskey, "Introducing huBERT," in XVII. Conference on Hungarian Computational Linguistics, 2019, pp. 3-14

[22] J. Ács, D. Lévai, D. M. Nemeskey, and A. Kornai, "Evaluating Contextualized Language Models for Hungarian," Feb. 2021, Accessed: Apr. 27, 2021 [Online] Available: http://arxiv.org/abs/2102.10848

[23] "HIL-SBERT." https://hilanco.github.io/models/sbert.html (accessed Jan. 20, 2022)

[24] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf., pp. 3982-3992, Aug. 2019, Accessed: Jul. 01, 2021 [Online] Available: http://arxiv.org/abs/1908.10084

[25] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," Feb. 2018, Accessed: Jul. 01, 2021 [Online] Available: http://arxiv.org/abs/1802.03426

[26] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering," J. Open Source Softw., Vol. 2, No. 11, p. 205, 2017, doi: 10.21105/joss.00205

[27] "MaartenGr/cTFIDF: Creating class-based TF-IDF matrices." https://github.com/MaartenGr/cTFIDF (accessed Jan. 20, 2022)

[28] "Hungarian Twitter Sentiment Corpus - opendata.hu." http://opendata.hu/dataset/hungarian-twitter-sentiment-corpus (accessed Jan. 20, 2022)

[29] "Precognox." https://www.precognox.com/ (accessed Jan. 20, 2022)

[30] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding," 7th Int. Conf. Learn. Represent. ICLR 2019, Apr. 2018, Accessed: Jul. 01, 2021 [Online] Available: http://arxiv.org/abs/1804.07461

[31] Ó. B. Fernandes et al., "Self-reported waiting times for outpatient health care services in Hungary: Results of a cross-sectional survey on a national representative sample," Int. J. Environ. Res. Public Health, Vol. 18, No. 5, pp. 1-15, Mar. 2021, doi: 10.3390/ijerph18052213

[32] A. J. Forster, H. J. Murff, J. F. Peterson, T. K. Gandhi, and D. W. Bates, "The incidence and severity of adverse events affecting patients after discharge from the hospital," Ann. Intern. Med., Vol. 138, No. 3, pp. 161-167, Feb. 2003, doi: 10.7326/0003-4819-138-3-200302040-00007

[33] M. Bahja and M. Lycett, "Identifying patient experience from online resources via sentiment analysis and topic modelling," Proc. - 3rd IEEE/ACM Int. Conf. Big Data Comput. Appl. Technol. BDCAT 2016, pp. 94-99, 2016, doi: 10.1145/3006299.3006335

[34] T. A. Ito, J. T. Larsen, N. K. Smith, and J. T. Cacioppo, "Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations," J. Pers. Soc. Psychol., Vol. 75, No. 4, pp. 887-900, 1998, doi: 10.1037//0022-3514.75.4.887

[35] "National Health Insurance Fund of Hungary – Inpatient Statistics." http://www.neak.gov.hu/virtualis_rovat/altfin_virt_dok2/besorolo/fekvo_stat (accessed Nov. 21, 2021)

[36] "Investigating Patients' Complaints." https://www.patientsrights.hu/investigating-patients-complaints.html (accessed May 31, 2021)

[37] T. Bourne et al., "Doctors' experiences and their perception of the most stressful aspects of complaints processes in the UK: An analysis of qualitative survey data," BMJ Open, Vol. 6, No. 7, pp. 1-10, 2016, doi: 10.1136/bmjopen-2016-011711

[38] Creative Commons Hungary, https://creativecommons.org/licenses/by-nc-nd/2.5/hu/ (accessed Jan 25, 2022)

Characteristics of Crisis Management Measures in the HR Area during the Pandemic in Hungary – Literature Review and Methodology

Krisztina Dajnoki, András István Kun

University of Debrecen Faculty of Economics and Business, Böszörményi út 138, H-4032 Debrecen, Hungary, e-mail: dajnoki.krisztina@econ.unideb.hu; kun.andras.istvan@econ.unideb.hu

József Poór

J. Selye University Faculty of Economics and Informatics, Bratislavská cesta 3322, 945 01 Komárno, Slovakia, e-mail: poorj@ujs.sk

Ákos Jarjabka

University of Pécs Faculty of Business and Economics, Rákóczi út 80, H-7622 Pécs, Hungary, e-mail: jarjabka.akos@ktk.pte.hu

Botond Géza Kálmán

Budapest Metropolitan University, Institute of Business
Nagy Lajos király útja 1-9, 1148 Budapest, eupemq@instructor.metropolitan.hu

Zsolt Sándor Kőmüves

Hungarian University of Agriculture and Life Sciences Kaposvár Campus, Guba Sándor utca 40, H-7400 Kaposvár, Hungary, e-mail: Komuves.Zsolt.Sandor@unimate.hu

Beáta Pató Szűcs

Eötvös Loránd University Faculty of Social Sciences, Egyetem tér 1-3, H-1053 Budapest, Hungary, e-mail: szucs.beata@sek.elte.hu

Katalin Szabó

Hungarian University of Agriculture and Life Sciences Gödöllő Campus, Páter Károly utca 1, H-2100 Gödöllő, Hungary, e-mail: Szabo.Katalin@uni-mate.hu

Szilvia Szabó

Budapest Metropolitan University Institute of Business, Nagy Lajos király útja 1-9, H-1148 Budapest, Hungary, e-mail: szszabo@metropolitan.hu

Zsuzsanna Szeiner

J. Selye University Faculty of Economics and Informatics, Bratislavská cesta 3322, 945 01 Komárno, Slovalia, e-mail: szeinerzs@ujs.sk

Arnold Tóth

Budapest Business School Faculty of Finance and Accountancy, Buzogány utca 10-12, H-1149 Budapest, Hungary, e-mail: toth.arnold@uni-bge.hu

Imola Csehné Papp

Eötvös Loránd University Faculty of Education and Psychology, Kazinczy u. 23-27, H-1075 Budapest, Hungary, e-mail: papp.imola@ppk.elte.hu

Abstract: In the present research we deal with the questions of what responses, solutions, and practices Hungarian organisations used to mitigate the impact of the pandemic and the changes caused by it, with particular attention to the area of Human Resource Management. Do they typically view the current crisis as a problem or an opportunity? This study has been divided into two papers. In this article, the authors review the international and Hungarian literature to reveal the most important negative and positive effects and aspects of epidemics from the theoretical as well as the practical aspect. The findings support that the typical measures and solutions of previous crises in history are partially or fully recurring. However, the current crisis has a very peculiar characteristic: the application of lockdowns. After the description of the general international context based on literature, we develop empirical hypotheses and describe the methodology of the conducted countrywide questionnaire survey. The survey results and the conclusions of the analysis are presented in a second article. In our opinion, the pandemic has reminded us of the extraordinary degree of interconnectedness of the world.

Thus, studies from individual countries – such as Hungary – and regions can provide important local insights in finding global solutions

Keywords: COVID-19; pandemic; crisis; human resource management; home office

1 Introduction

The COVID-19 pandemic had drastic consequences for the entire society. It can be generally stated that the poor and vulnerable people in the world [1, 2], as well as those who are more exposed (young people, migrants, and the elderly) have become even more so [3, 4]. For the purpose of stopping the spread of the virus, the governments of numerous countries in the world ordered unprecedented intervention, which in most cases resulted in the lockdowns of entire regions and countries [5, 6, 7]. The suspension of economic activities had serious consequences for employment [8, 9] and the income of employees as well as the revenue of companies. Drastic changes became necessary on every level of society to protect health and security [10]. Companies could not stay out of the pressure of this fast reaction either, they had to retain their most important resource, human capital. It represented a particularly great challenge how to assist their employees to adapt and overcome the changes that occurred in workplaces and social life. For example, how to create a home office without violating the interests of either the employer or the employee? How can the separation between work and private life be ensured [11, 12]? In the present research, we deal with the questions of what responses, solutions, and practices organisations used against the appearance of the pandemic and the changes caused by it, in general, and with particular attention to the area of Human Resource Management (HRM). Did they view the current crisis as a problem, challenge, or possibly as an opportunity? The results and conclusions of our work are contained in our second article. According to Budhwar and Cumming [13], the COVID-19 crisis has drawn attention to the importance of an international perspective. As we wrote earlier, in their opinion, the pandemic reminded us of the extraordinary degree of interconnectedness of the world. Thus, studies from individual countries and regions can provide important local insights, e.g. in relation to the organizational HRM management of the epidemic, they can also help us find global solutions [14].

As we presented in the abstract, the main contributions of the first part of our study are the empirical hypotheses derived from literature analysis, and the description of the methodology for testing these hypotheses. The results will be reported and the conclusions will be discussed in the second article.

2 Literature Review

In human resource management, which was considered one of the most problematic areas of the management of East-Central-European companies in the first decade after the transition [15, 16, 17], *"by the beginning of the new millennium, considerable positive changes had occurred at Hungarian companies"* [18]. In Hungary the total number of employees is 4.5 million, and 13.33% [19] of them are employed by multinational companies that settled in the country, where the modern HRM principles and methods of parent companies were quickly adopted by Hungarian companies into their everyday work [20]. Because of rising market competition, an increasing number of *Hungarian (domestically) owned* organisations apply innovative HR knowledge and professionals with this type of knowledge [21]. Initially, the expert supporters of the development that commenced in the area of human resource management were large – also international – *HR consultancy companies*, which were later joined by more and more Hungarian consultancy companies [22]. Nowadays, in the Bologna system bachelor's and master's degree programmes, an increasing number of colleges and universities offer high-quality HRM programmes. Before the breakout of the coronavirus epidemic, in Hungary – similar to other Eastern-European countries [23, 24, 25, 26] – substantial workforce shortage developed in certain sectors [27], which to a certain degree 'was alleviated' during the epidemic. But during the second wave, the first signs of the previous situation started to show. By today the rate of employment has reached and exceeded the 70% level [28].

First wave: On March 11, 2020, the Hungarian government declared a state of emergency for the entire territory of the country. The repayment of principal and interest rate was suspended until the end of the year 2020. In the sectors with serious problems (tourism, hospitality, entertainment, gambling, film, performing arts, event management, and sports); by 30 June, employers were released from the obligation to pay public charges [29]. According to analysts, the first wave of the pandemic affected the Hungarian labour market more severely than it appears from official statistics [28, 30], partially, for instance because workhour reductions are not indicated in the data and partially because of 'waiting it out' and informal solutions that are characteristic of both sides [31].

Second wave: From September 2020, the reintroduction of emergency measures and more rigorous mask wearing obligations occurred. The government extended the credit moratorium for 6 months. Administrative burdens were further reduced, from the second half of 2020 the tax office prepared a draft VAT return for each company. Business support program for enterprises: up to half a million forints per job seeker was introduced. Employment is picking up again, and the wage support programs established in the economic protection action plan still protected for more than a quarter of the total number of 280,000 beneficiaries. From

November, the attendance regulations for mass events as well as the rules for wearing a mask were tightened again.

The effect of the COVID-19 pandemic caused broad and great challenges to employees, organisations, communities, nations as well as the entire world [32]. During previous economic crises the lockdown mode had never been applied [33]. However, the world did know its effects in connection with the Nipah virus epidemic in Vietnam, which was more dangerous than COVID [34]. The currently applied methods, government intervention methods, and multilateral solutions greatly differ from the crisis management measures applied in 2008 [35]. Unlike previous crises [36, 37], coronavirus led to a recession, not because of overproduction or loss of confidence, it rather endangered human resources that represent the foundation for the growth of modern economies [38, 39, 40, 41, 42]. With this, the protection of human resources and an increase in their significance gained new momentum. The study by [43] points out that during an economic downturn, those who are still employed may experience workplace uncertainty caused by reductions in salary and work hours, deteriorating physical and mental health, long forced vacations, and challenging conditions.

Worldwide COVID-19 has a yet unpredictable effect on social and economic actors, and here the questions of work organisation during the global crisis (and after it) can be identified as a serious research trend. Specifically, it is a shared interest of employers and employees to achieve undisturbed business processes during the pandemic, for the employers to continually employ the employees who thus receive their salary, to avoid mass layoffs and downsizing [44]. The legal-political reactions [45, 46], restrictive measures, and limitations of mobility introduced as a consequence of the pandemic considerably transformed the HR practices generally applied in the labour market. The majority of organisations took the following important measures as a response: stopped business travel, company events, provided disinfectants and masks at workplaces, furthermore permitted or ordered work-from-home and flexible work performance [47]. During decision making processes management had to consider two important issues: 1) Which things are the most important to employees, and along with this 2) What is important for HR?

From the aspect of companies one of the greatest problems was retaining liquidity and within that particularly paying salaries. Despite this fact, more than half of all companies considered it extraordinarily important to retain their pre-pandemic headcount, either by work organisation or workhour and pay reductions, or by remote work, and in the worst-case scenario by forced unpaid vacations instead of layoffs [48]. Based on [49], the social distancing effect of the pandemic weighs more heavily on working women, the elderly, and immigrants. The loss of jobs and the reduction of income can have an effect, particularly on temporary workers and those with lower qualifications [50].

From the aspect of employees, the most important factor is the loss of income or the fear of it. Financial troubles among the population of all member states worsened, and a loss of income was felt by over 30% of respondents in 21 member states, in other countries people were more concerned about being forced to spend savings [51]. The crisis intensified inequalities (e.g. jobs, salaries, and workplace opportunities), especially in the case of young people and less qualified employees [37, 50].

The other challenge was creating a balance between work and private life. Researchers discovered a number of challenges in connection with this (e.g. increased stress, loneliness, increased overtime at home, merging of worktime and private life, etc.) [52, 53]. Although according to the latest forecasts, instead of the creation of a work-private life balance, integrated well-being design is emerging. According to the experts of [54], these two areas of life can only function well if they are integrated. During the pandemic, we have witnessed phenomena such as the development of social isolation, segregation, uncertainty, and sense of fear. In this situation, organisations must be more flexible and innovative in unexpected HR related issues that affect employees. The coronavirus has forced organisations to introduce new and radical methods in the area of work performance, these processes had started earlier, but the pandemic greatly accelerated them [55]. According to the above-mentioned research by [54], the future of work is influenced by three factors that are closely related to each other. The first is the type of work, as a result of increasing automation certain job positions are eliminated or transformed and in parallel with this new ones are created. The second is the issue of the workforce: who will perform the work? New combined talent models and diverse workforce-stocks are established. The third factor is the site of work performance, the experience of the past year and a half definitely established the basis for the long-term application of flexible work strategies.

The COVID-19 pandemic moved millions of employees from offices to their homes worldwide, for the purpose of isolating themselves. During the quarantine a considerable number of articles were published in the press about the difficulties of remote work, and the suggestions of business consultants regarding how to organise work and cooperation [56, 57], through the recommendations of HR consultants regarding how to maintain the productivity of employees [58, 59], to comments by psychologist regarding the balance of work and private life in a situation where work and private life go on simultaneously without leaving home [60]. At the same time, it is important to emphasise that the pandemic-related advice and suggestions of experts and consultants were based on the knowledge that had been accumulated before the pandemic.

Based on [61], employees are more concerned about their financial security, and personal and family health than the financial situation and possible instability of their employers, thus the HR strategy must consider the concerns of employees. According to Hungarian research experience [62], the greatest challenge to

employers is effective internal communication, informing, and maintaining contact. This is followed by the difficulty of workforce retention, the introduction/operation of a home office and remote work, as well as compliance with health protection and hygiene regulations. Critical points are maintaining motivation, wage management, panic management, and the development of digital processes. In the future, they must strive to develop strategies where practical application contributes to balancing the relationship between the workforce and the changed labour market conditions, and it is necessary to facilitate the development of effective learning methods by utilising the opportunities provided by digital transformation. The above train of thought was supplemented by [63]; the engine of organisational development and HR are creativity and innovation, which human resource possesses as human capital, thus its role is expected to grow exponentially in the near future. By [64], despite the current era of innovation and automation, the human factor is irreplaceable.

As it can be seen from the above, the coronavirus epidemic also plays a significant role in the HR activities of organisations. According to [65], the biggest challenge for HR professionals is how to contribute to the survival of companies in this challenging situation. In order to adapt to rapidly changing market needs, they need to incorporate the tools of agile methods into their operations. Indeed, the restrictive measures and mobility restrictions implemented as a result of the pandemic have significantly transformed HR practices that have become commonplace in the labour market.

The state supplements the role of HR with social and economic assistance measures – including fiscal incentives and the expansion of social safety nets – which are essential to prevent the further rippling of the effects of the pandemic [66]. Of course, money is required to solve the economic problems caused by the pandemic, but that should not be invested into production, it rather should be spent on the protection and development of human capital. It is not a coincidence that government-level measures primarily serve the protection of health and jobs [67, 68]. On the company level, these same measures are in the range of tasks of HR. So far, Hungarian experts have processed the experiences of the pandemic's first and second wave. Broad-ranging company-level surveys were conducted about the issue during the spring and autumn waves as well [69]. Based on the research reports it can be stated that on top of the usual measures new solutions also appeared, moreover were given central roles, which HR professionals expressly adapted to the pandemic as the root cause.

3 Hypotheses

We would like to contribute to the above-introduced literature in the area of HR crisis management by testing the following two hypotheses empirically. The first one is derived from the findings of [70, 71]:

H1: The typically occurring crisis management measures [70] in the area of HRM caused by the effects of the pandemic are primarily related to headcount management/workforce requirements (hiring freeze, downsizing, downsizing of temporary staff, reducing labour requirements by automation/technical solutions and the innovative handling of work hours like reduction of working hours, enabling/directing home offices, elaboration/re-planning of replacement plans) [71].

Our second hypothesis is built on the previous research by [72, 73, 74, 75, 76, 55]:

H2: In the sample, the most frequently applied HR crisis management measures are correlated with the organisation's revenue and headcount: Enabling/directing home offices [72], New occupational health and safety measures [73], hiring freeze and downsizing are considerably more frequent at larger companies than at smaller ones [75, 76]. We supplemented the hypothesis H2 with sub-hypotheses.

H2a: In the sample, the most frequently applied HR crisis management measures are correlated with the organisation's two main indicators: revenue and headcount.

H2b: Enabling/directing home offices, new occupational health and safety measures, and reducing the risks of the pandemic HR crisis management measures [55] are correlated with the organisation's two main indicators: revenue and headcount [74].

H2c: The "No task" attitude is mainly typical at organisations that are smaller according to revenue and headcount.

4 Material and Methodology

We tested our above-specified hypotheses on the databases of the first two phases of our Covid-19 HR research. The research team was established with the joint effort of the professors of 14 Hungarian universities and 1 Slovakian university. The nationwide online survey started at the beginning of May 2020, a few months after the first European infections caused by COVID-19. To examine the topic, we conducted benchmark-type research, which allows us to monitor and compare the changes in the selected indicators over time. Such studies are typically run regularly (monthly, quarterly, annually) and evaluate how the experiences with a

particular issue change over time [77, 78]. In the emergency caused by the pandemic, employers had to face daily changes, an uncertain, unpredictable situation. Benchmarking allows us to track the progress of the changes and revisions, therefore we conducted our research several times in a relatively short period, so that the changes in HR solutions during the coronavirus crisis could be monitored.

1) The first phase of the data collection was conducted between 12 June and 31 July 2020 with a questionnaire survey, where the number of respondent organisations was 508.
2) The second phase of data collection was conducted between 1 August and 15 November 2020 by an online questionnaire survey, where we received 1,041 useful responses from various organisations.

The survey is fundamentally ex-post [79] type, thus it examined the effect of the crisis caused by Covid-19 on the human resource management practices in the Hungarian company/institutional sector based on opinions and factual data related to the surveyed time period. Figure 1 describes the research model in connection with the hypothesis.

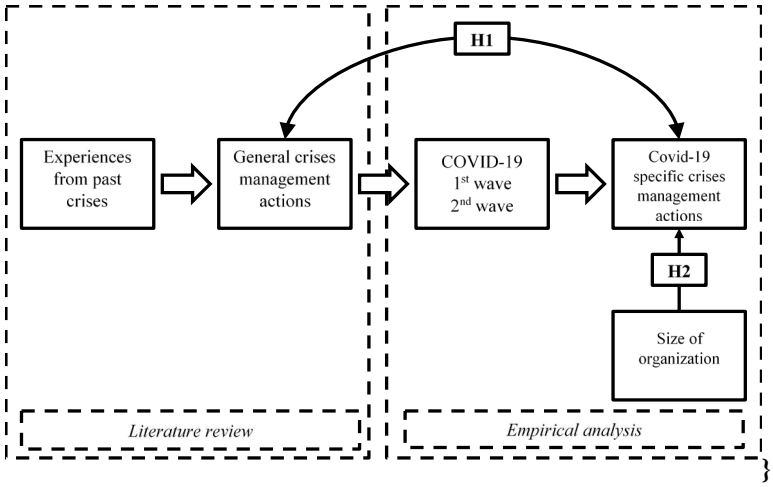


Figure 1
Research concept

The results of the two phases of the research are not directly comparable, since the range of respondents did not remain unchanged, thus in the present study we analyse them separately.

Figure 2 summarises the distribution of respondents according to the form of ownership. In the two phases, a considerable portion of the respondents were Hungarian private organisations (58.7% and 46.9%, respectively) and approximately one-quarter of them were in foreign or mixed ownership.

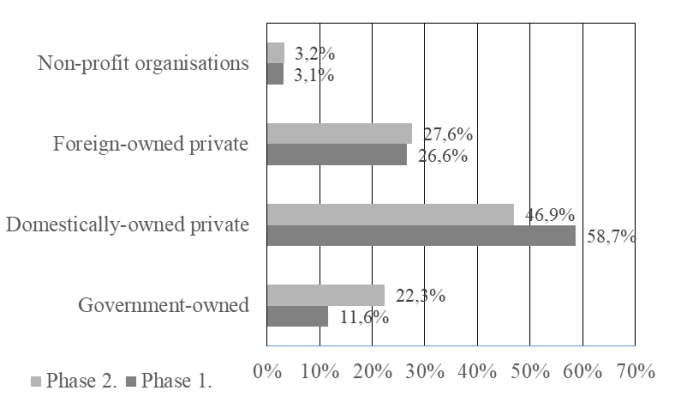


Figure 2
Distribution of respondents according to form of ownership during the first and second phase

We conducted the examination of organisation size based on two aspects, employee headcount and revenue (according to the practice of the [80], among others). For the interpretation of the below-presented data and results, we describe some Hungarian data as a basis for comparison.

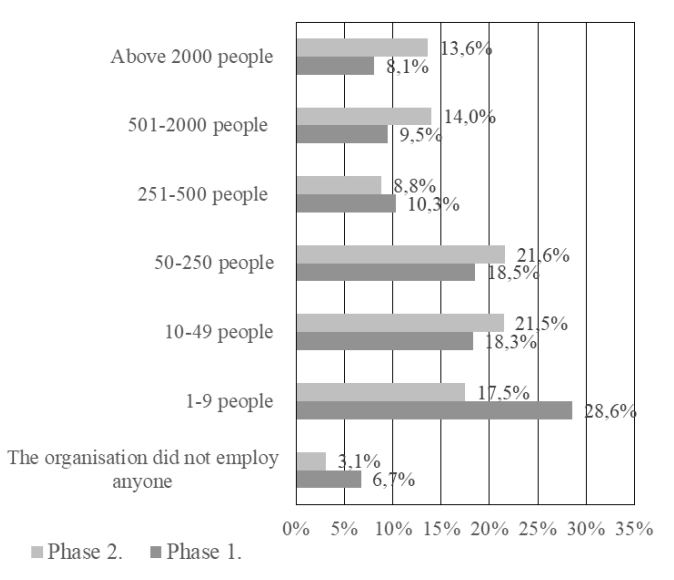


Figure 3
Distribution of respondents according to headcount during the first and second phase

Based on the 2019 data of [19], regarding the number of enterprises, the overwhelming majority (91.4%) have an employee headcount of 1-10 people. 3.8% of all enterprises have an employee headcount of 10-49 people. Employee headcount of 50-249 people is at 0.6% of companies, and over 250 people at 0.1%, while at 4% of registered companies, there are no employees at all. In 2020, in Hungary, the number of budgetary organisations and organisations managed according to budgetary rules was around 12,800, which is 0.7% of registered business organisations [81]. Based on employee headcount, the highest percentage of respondents (in the first phase 72.1%, in the second phase 63.6%) were in the SME category (Figure 3).

Another widely applied indicator in categorising organisation size (see [80]) describes samples according to revenue (Figure 4). Based on KSH data, in 2019 the average revenue of business organisations was around HUF 141.6 million [19]. From among the respondents in the first phase, 37.4% fell in the category under this revenue, while in the second phase 32.4%, thus the proportion of respondents with lower-than-average revenue decreased by 5% in the sample. The proportion of respondents with higher than the national average revenue was 45.5% at the first data collection, and it rose to 50.9% by the second phase (i.e. by 5.4%).

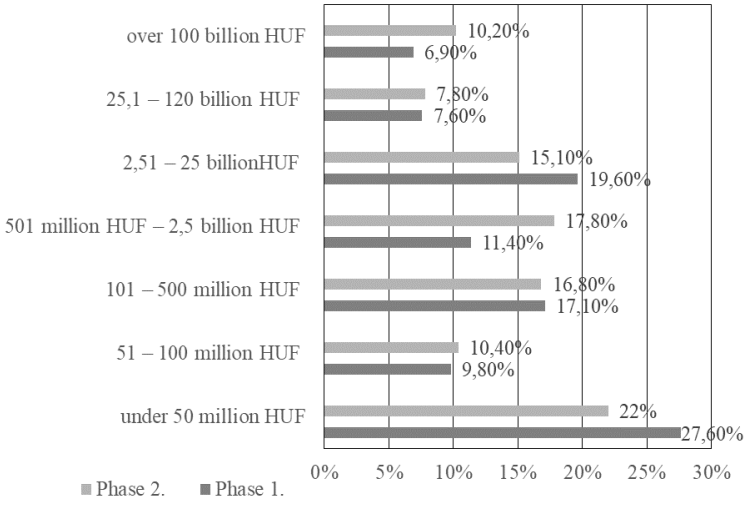


Figure 4

Distribution of respondents according to revenue during the first and second phase

* 1 EUR = 351.1 HUF [80]

The same questionnaire was used during both phases of the survey (in original language: http://limesurvey.szie.hu/index.php/44678?lang=hu; the authors will provide the English translation electronically) and the questions highlighted for the present study *("Annual budget/revenue in 2019", "Average number of*

employees in 2019", "Please indicate if your firm practiced the following crisis management HR measures for your organisation") were unchanged. The applied questionnaire were discussed with academic experts and then with practicing HR executives to ensure understanding and validity. The instrument was finalized after several rounds of tryout and correction.

We have conducted the following statistical methods to test our hypotheses.

We conduct the examination of the H1 hypothesis by the analysis of frequencies and relative frequencies of respones to questions about the use of 19 HR crisis management measures in both waves (separately).

We inspect the H2 hypothesis through its sub-hypotheses. For the examination of the H2a, H2b, and H2c hypotheses – since revenue (I2), headcount (I3), and the question related to HR measures (I4) can be coded as sequential variables –, the analyses were performed with Spearman's rank correlation (which does not require ratio-scale variables or normality). The examination tests if there is a correlation between size variables (headcount or revenue) and the use of the 19 HR crisis management measures. The analyses were performed with pairwise method, thus the sample size differ by the tested variable-pairs. Samples from the two surveys are examined separately.

In the first round we examine the correlations between the HR measure and the revenue variable as well as the headcount variable, then to examine H2a we checked if among the measures showing a significant correlation, there are some that were found to be the most typical at the examination of H1, for the examination of H2b we checked the significance of the "Enabling/directing home offices", "New occupational health and safety measures", and "Reducing the risks of the pandemic through training" questions, while for H2c we inspected if there is a correlation between the "No tasks" measure and the revenue variable as well as the headcount variable, and if yes, whether the coefficient is negative.

We also analysed the H2c sub-hypotheses with the chi-square test (by calculating Cramér's *V*). For this we transformed the four-value "No task" variable into three dummy variables:

- *No task at all*: its value is 1 if the respondent somewhat or more agreed that there is no task, otherwise 0.
- *Medium level no task*: if the respondent intermediately or more agreed that there is no task, otherwise 0.
- *Very much no task*: if the respondent strongly agreed that there is no task, otherwise 0.

In this way, we can determine if we can find a significant divergence between each headcount and revenue category as well as combined categories (e.g. we can create two headcount categories of 0-49 people and over 50 people). The Results chapter – in the second part of our paper – describes the formation of categories

because that is based on the preceding survey results, thus it would be early to show them here.

As it was mentioned above since the two samples are structured differently, and there may be overlaps between respondents, they are always analysed separately. This is otherwise also justified by the time that passed between the data collection and the different pandemic situations.

5 Summary of the First Part

In the presented part of our study, we have reviewd the literature background of the possible negative and positive effects and aspects of the COVID-19 pandemic on HRM practice. As the main output of this paper, based on this literture analysis we have drawn two empirical hypotheses and described the methodology to test them.

The presentation and discussion of the empirical results are presented in the second part of this study, which is titled *"Characteristics of Crisis Management Measures in the HR Area during the Pandemic in Hungary – Reflected in a Nationwide Empirical Research"*. In that article, we report the results of testing the hypotheses, we draw our conclusions and discuss limitations of our research, as well as possible future research directions.

References

[1] Ch. Lakner: How Much Does Reducing Inequality Matter for Global Poverty?, working paper, World Bank Group, Washington, 1 June 2020

[2] N. Madarász, B. Pásztor, K. Lazányi: Individual decisions during the coronavirus. In: Keszthelyi, A. et al (eds.) 18th International Conference on Management, Enterprise, Benchmarking. Proceedings (MEB 2020) Budapest, Hungary, Óbuda Universty, 2020, 118-126

[3] Ferenci T.: Different approaches to quantify years of life lost from COVID-19, Eur J Epidemiol 36, 2021, pp. 589-597

[4] Z. D. Paget: "COVID-19 and its impact on society", available at: https://www.openaccessgovernment.org/covid-19-and-its-impact-on-society/110928/ (accessed 26 June 2021)

[5] IMF "Policy Responses to COVID 19", working paper, International Monetary Fund, Washington, 1 July 2021

[6] X. Chen, Z. Qiu: COVID-19: Government interventions and the economy, available at: https://voxeu.org/article/government-interventions-covid-19-and-economy (accessed 26 June 2021)

[7] T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, H. Tatlow: A global panel database of pandemic policies (Oxford Government Response Tracker), Nature Human Behaviour, Vol. 5, No. 4, pp. 529-538 (2021)

[8] D. Gros, A. Ounnas: Labour market responses to the Covid-19 crises in the United State and Europe, working paper WD2021-01, Centre for European Policy Studies, Brussels, April 2021

[9] Röst G, Bartha F. A., Bogya N., Boldog P., Dénes A., Ferenci T., Horváth K. J., Juhász A., Nagy C., Tekeli T., Vizi Z., Oroszi B.: Early Phase of the COVID-19 Outbreak in Hungary and Post-Lockdown Scenarios. Viruses. 12(7):708. 2021

[10] D. W. Eggers, M. Flynn, J. O'Leary, B. Chew: Governments' response to Covid-19 From pandemic crisis to a better future, available at: https://www2.deloitte.com/us/en/insights/economy/covid-19/governments-respond-to-covid-19.html (accessed 1 June 2021)

[11] N. Chawla, R. MacGowan, A. Gabriel, N. P. Podsakoff: Unplugging or staying connected? Examining the nature, antecedents, and consequences of profiles of daily recovery experiences Journal of Applied Psychology, 2019, Vol. 105, No. 1, 19

[12] J. B. Carnevale, I. Hatak: Employee adjustment and well-being in the era of COVID-19: Implications for human resource management, Journal of Business Research, 2020, Vol. 116, 183-187

[13] P. Budhwar, P., D. Cumming: New Directions in Management Research and Communication: Lessons from the COVID-19 Pandemic, 2020, British Journal of Management, Vol. 31, 3, 441-443, https://doi.org/10.1111/1467-8551.12426

[14] N. Sipos, Á. Jarjabka, G. Kuráth, T. Venczel-Szakó: Higher education in the grip of COVID-19: 10 years in 10 days? Quick report on the effects of the digital transition in work at the University of Pécs, 2020, Civil Review, special issue 1, 71-90

[15] J. M. Morley, N. Heraty, S. Michailova: Studying human resource management in the international context: The case of Central and Eastern Europe, Managing Human Resources in Central and Eastern Europe. London, Routledge, 2009, 1-24

[16] Halmos K.: The effect of FDI, exports and GDP on income inequality in 15 Eastern European countries. Acta Polytechnica Hungarica, 2011, Vol. 8, No. 1, 123-136

[17] J. M. Morley, D. Minbaeva, S. Michailova: HRM in transition in the transition states of Central and Eastern Europe and the former Soviet Union, Brewster, Ch. et al. (Eds.) Handbook of Research on Comparative

Human Resource Management, Edward Elgar Publishing, London, UK, 2018, 469-487

[18] Karoliny Ms.: Overview of human resource management (In Hungarian), 23-60, In Karoliny Ms. and Poór J. Handbook of Human Resource Management Budapest: Wolters Kluwer, 2017

[19] KSH Enterprise performance indicators by headcount category (in Hungarian), available at: http://www.ksh.hu/docs/hun/xstadat/xstadat_eves/i_qta001.html (accessed 1 January 2021)

[20] Bácsi K., Szőtsné Kovács K., Takács S., Toarniczky A.: Human Resource Management, leadership and competitiveness (in Hungarian), working paper, Corvinus University, Budapest, Hungary, 31 August 2007

[21] Bakacsi Gy. Takács S., Bokor A., Császár Cs., Gelei A., Kováts Klaudia: Strategic Human Resource Management (in Hungarian) KJK-KERSZÖV Publishing House, Budapest, Hungary, 2006

[22] Consultancy.eu: Consulting industry of Hungary breaks through €400 million mark, available at: https://www.consultancy.eu/news/1803/consulting-industry-of-hungary-breaks-through-400-million-mark (accessed 7 July 2021)

[23] Z. B. Schwidrowski, W. Li, T. Yousef: Skill and labor market outcomes in Central Europe. Economic Systems, Vol. 33, No. 1, 45-59

[24] Takács M.: Multilevel fuzzy approach to the risk and disaster management. Acta Polytechnica Hungarica, Vol. 7, No. 4, 91-102

[25] C. Patricolo: Czech Republic looks east to ease labour shortages, available at: https://emerging-europe.com/news/czech-republic-looks-east-to-ease-labour-shortages/ (accessed 1 May 2020)

[26] J. Łukomska-Szarek, A. Martynko, Z. Warzecha: Management under Crisis Conditions – the Impact of the COVID-19 Pandemic on the Formation of Respondents' Opinions within the e-commerce Market, in Poland. Acta Polytechnica Hungarica, 2021, Vol. 18, No. 11, 251-267

[27] PWC: Central and Eastern Europe Private Business Survey 2019, available at: https://www.pwc.com/hu/hu/kiadvanyok/assets/pdf/PwC-PBS-CEE-Report-2019-final.pdf (accessed: 1 June 2020)

[28] Bakó T., Lakatos J.: The Hungarian labour market in 2019–2020. Fazekas, K. et al. (Eds.), The Hungarian Labour Market (in Hungarian), Centre for Economic and Regional Studies Institute of Economics, Budapest, Hungary, 2021, 15-34

[29] Kaszás F.: First Wave vs. Second Wave: When Did the Gov't Implement Stricter Restrictions?, Hungary Today, 13 November 2020, available at https://hungarytoday.hu/first-wave-vs-second-wave-when-did-the-govt-implement-stricter-restrictions/ (accessed 3 July 2021)

[30] Köllő J., Reizer B.: The impact of the first wave of the coronavirus epidemic on employment and corporate sales (in Hungarian), Hungarian Economic Review, 2021, Vol. 68, No. 4, 345-374

[31] Palócz É., Matheika Z.: Dilemmas on the economic effects of the Covid-19 crisis in Hungary (in Hungarian), Kolosi, T., et al. (Eds.) Társadalmi Riport TÁRKI, Budapest, Hungary. 2020, 573-590

[32] J. Li, R. Ghosh, S. Nachmias: In a time of COVID-19 pandemic, stay healthy, connected, productive, and learning: words from the editorial team of HRDI. Human Resource Development International 2020, Vol. 23, No. 3, 199-207, DOI: 10.1080/13678868.2020.1752493

[33] Lazáry J.: Acute effects of the COVID-19 pandemic on mental status in the first phase of the epidemic, Neuropsychopharmacologia Hungarica 2020, Vol. 22, No. 2, 172-177

[34] V. Thomas: Containing a deadly virus: Lessons from the Nipah outbreak in India. Brookings, https://www.brookings.edu/blog/future-development/2018/07/23/containing-a-deadly-virus-lessons-from-the-nipah-outbreak-in-india/ (accessed: 12 February 2021)

[35] M. E. Strauss-Kahn: Can we compare COVID-19 and 2008 crises?, available at: https://www.atlanticcouncil.org/blogs/new-atlanticist/can-we-compare-the-covid-19-and-2008-crises/ (accessed 17 Febr 2021)

[36] A. S. Blinder, M. Zandi: The Financial Crisis: Lessons for the Next One. CBPP, available at: https://www.cbpp.org/research/economy/the-financial-crisis-lessons-for-the-next-one (accessed 4 March 2021)

[37] A. Adams-Prassl, T. Boneva, M. Golin, C. Rauh: Inequality in the impact of the coronavirus shock: Evidence from real-time surveys, Journal of Public Economics, Vol. 189, https://doi.org/10.1016/j.jpubeco.2020.104245

[38] R. W Fogel: Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy, The American Economic Review 1994, Vol. 84, No. 3, 369-395

[39] T. W. Schultz: Investment in human capital, The American Economic Review 1961, Vol. 51, No. 1, pp. 1-17

[40] T. W. Schultz: The Economics of Being Poor, Bulletin of the Atomic Scientists 1980, Vol. 36, No. 9, 32-37

[41] D. Torrington, L. Hall, S. Taylor, C. Atkinson: Human Resource Management. Harlow: Pearson Education Limited, 2014

[42] D. Ulrich, J.H. Dulebohn: Are we there yet? What's next for HR? Human Resource Management Review 2015, Vol. 17, No. 1, 1-9

[43] M. R. Frone: What happened to the employed during the Great Recession? A U.S. Population study of net change in employee insecurity, health, and

organizational commitment, Journal of Vocational Behaviour 2018, Vol. 107, 246-260

[44] Csiszárik-Kocsir Á., Garai-Fodor M., Varga, J.: What has Become Important during the Pandemic? – Reassessing Preferences and Purchasing Habits as an Aftermath of the Coronavirus Epidemic through the Eyes of Different Generations. Acta Polytechnica Hungarica 2021, Vol. 18, No. 11, 49-74

[45] Kun A.: Hungary – The impact of the labour law measures taken by the authorities: reflections one year after the official recognition of Covid-19 as a pandemic (in Hungarian), working paper, Budapest, Hungary, 1 March 2021

[46] Váradi B.: Policy responses to the coronavirus pandemic in Hungary during the first half of 2020, Fazekas, K. et al. (Eds.) The Hungarian Labour Market 2020, Centre For Economic and Regional Studies Institute of Economics 2021, Budapest, Hungary, 204-210

[47] Vörös M., Fűrész D. I.: Empirical study of the effectiveness of part-time employment (in Hungarian), Hungarian Economic Review 2021, Vol. 68, No. 2, 178-204

[48] MNB: The MNB's corporate survey examining the economic effects of the coronavirus epidemic (in Hungarian), available at: https://www.mnb.hu/sajtoszoba/sajtokozlemenyek/2020-evi-sajtokozlemenyek/az-mnb-koronavirus-jarvany-gazdasagi-hatasait-vizsgalo-vallalati-felmeresenek-eredmenyei (accessed 20 February 2021)

[49] Csehné Papp I., Varga, E.: Epidemic phenomenon in a labour market context. Controller Info 2021, Vol. 9, No. 1, pp. 51-55

[50] Bagó J.: Pandemia and work (in Hungarian), New Hungarian Labour Review 2020, Vol. 1, No. 3, pp. 14-25

[51] European Parliament: Uncertainty | EU | Hope – Public opinion in times of COVID-19, 1-23, available at: https://www.europarl.europa.eu/resources/library/media/20200713RES83231/20200713RES83231.pdf (accessed: 4 March 2021)

[52] P. Bilge, C. A. Alkan, R. Ağanoğlu: Managing work-life balance during the Covid-19 crisis – A survey with 1500+ participants worldwide. Berlin: Technical University, 2020

[53] Szellő J.: Pandemic and the Global Labour Market (in Hungarian), New Hungarian Labor Review 2021, Vol. 1, No. 4, pp. 17-22

[54] Deloitte: The social enterprise in a world disrupted. Leading the shift from survive to thrive – 2021 Deloitte Global Human Capital Trends, available at: https://www2.deloitte.com/content/dam/insights/us/articles/6935_2021-HC-Trends/di_human-capital-trends.pdf (accessed 4 March 2021)

[55] ILO: Guidelines on occupational safety and health management systems, ILO-OSH 2001. Geneva: International Labor Organization. Second edition: Guidelines on occupational safety and health management systems, ILO-OSH 2001 (accessed: 3 July, 2021)

[56] T. Kawashima, S. Nomura,Y. Tanoue, D. Yoneoka, A. Eguchi, S. Shi H. Miyata: The relationship between fever rate and telework implementation as a social distancing measure against the COVID-19 pandemic in Japan. Public Health Vol. 192, No. 9, pp. 12-14, doi: 10.1016/j.puhe.2020.05.018

[57] A. Belzunegui-Eraso, A. Erro-Garcés: Teleworking in the Context of the COVID-19 Crisis, Sustainability 2020, Vol. 12, No. 9 (special issue) 1-18

[58] S. Raghuram, N. S. Hill, J. L. Gibbs, L. M. Maruping: Virtual Work: Bridging Research, Academy of Management Annals, Vol. 13, No. 1, 1-34, doi: 10.5465/annals.2017.0020

[59] H. Bouziri, D. R. M. Smith, A. Descatha, W. Dab, K. Jean: Working from home in the time of COVID-19: How to best preserve occupational health? Occup. Environ. Med., Vol. 77, pp. 509-510, doi: 10.1136/oemed-2020-106599

[60] K. Čulo: Virtual Organization – The Future Has Already Begun, Media, Culture and Public Relations 2016, Vol. 7, No. 1, pp. 35-42

[61] W. Al Mala: How COVID-19 Changes the HRM Practices (Adapting One HR Strategy May Not Fit to All), 2SRN Electronic Journal. http://dx.doi.org/10.2139/ssrn.3736719

[62] HR Portal: The companies had a ready-made action plan for the second wave. HR, available at: https://www.hrportal.hu/hr/kesz-cselekvesi-tervvel-rendelkeztek-a-vallalatok-a-masodik-hullamra-20210106.html (accessed 20 December 2020)

[63] Hegedűs H.: Review of the publication "The World of Work at the Beginning of the 21st Century. Handbook on Employment Policy and the Labor Market" (in Hungarian), Miltary Science, 2018, Vol. 28, E-Issue pp. 185-187

[64] K. Lazányi, Pál. Fehér-Polgár, I. Vida: Identification of the most important European productivity factors through the dimension reduction. Economic and Managerial Spectrum, 2020. 14:1,77-86

[65] A. Gulo: Transformational Human Resources Through Remote Working and Upskilling for Digital PT Eigerindo MTI.SSRN; 2021 unpublished manuscript, Europe PMC doi: 10.2139/ssrn.3815399

[66] D. Laborde: COVID-19 risks to global food security, Science, 2020, Vol. 369, No. 6503, 500-502

[67] MTI: Almost 10,000 companies have benefited from job protection wage subsidies (In Hungarian), HR-portál. https://www.hrportal.hu/c/majdnem-

10-ezer-vallalat-vette-igenybe-a-munkahelyvedelmi-bertamogatast-20200526.html (accessed: 20 February, 2021)

[68] MTI: Thousands of applications for job protection wage subsidies are received every day. HR Portál. https://www.hrportal.hu/hr/naponta-tobb-ezer-kerelem-erkezik-a-munkahelyvedelmi-bertamogatasra-20200619.html (accessed: 20 February, 2021)

[69] Balogh, G. et al.: Objectives, methods and first experiences of the Hungarian national Korona HR research project (in Hungarian) New Hungarian Labour Review, 2020, Vol. 2, No. 4

[70] G. D. Collings, J. McMackin, A. J. Nyberg, P. M. Wright: Strategic Human Resource Management and COVID-19: Emerging Challenges and Research Opportunities, 2021, Vol. 58, No. 5, 1378-1382

[71] S. Hamouche: Human resource management and the COVID-19 crisis: implications, challenges, opportunities, and future organizational directions, Journal of Management & Organization, 1-16, doi:10.1017/jmo.2021.15

[72] N. Lewis: HR Managers Rethink Their Role During the Coronavirus Pandemic available at: https://www.shrm.org/hr-today/news/hr-news/pages/hr-managers-rethink-their-work-coronavirus-pandemic.aspx (accessed 1 June 2021)

[73] F. Weidemeyer: COVID-19 crisis management: ten better questions to ask, available at: https://www.ey.com/en_gl/covid-19/covid-19-crisis-management-essential-ten-better-questions-to-ask (accessed 1 June 2021)

[74] N. Chanana, Sangeta: Employee engagement practices during COVID-19 lockdown, academic paper, J. Public Affairs □2020;e2508□ Whiley, Hoboken, New Jersey, Oct 1 2020, doi:10.1002/pa.2508

[75] IFC: How Firms are Responding and Adapting During COVID-19 and Recovery opportunities for accelerated inclusion in emerging markets, working paper, International Finance Corporation (IFC, a member of World Bank Group), Washington

[76] OECD: OECD Economic Outlook Vol. 2020 Issue 1, No. 107, https://doi.org/10.1787/0d1d1e2e-en

[77] B. Karlöf: Benchmarking, in Bidgoli, H. (ed.) Encyclopedia of Information Systems. pp. 65-80, ISBN 9780122272400, Elsevier, doi: https://doi.org/10.1016/B0-12-227240-4/00005-8

[78] S. Mannan (ed.): Benchmarking in the Process Industry, in Lees' Loss Prevention in the Process Industries (Fourth Edition), Butterworth-Heinemann, 2012, 2484-2491, ISBN 9780123971890, doi: https://doi.org/10.1016/B978-0-12-397189-0.00038-0

[79] Usunier, J-C., H. van Herk, J. A. Lee: International and Cross-Cultural Business Research. SAGE Publications Ltd., 2017, ISBN: 9781473975897

[80] European Comission: Economic Forecast: A deep and uneven recession, an uncertain recovery. available at: https://ec.europa.eu/info/business-economy-euro/economic-performance-and-forecasts/economic-forecasts/spring-2020-economic-forecast-deep-and-uneven-recession-uncertain-recovery_fr (accessed 12 June 2021)

[81] KSH: STADAT – 3.2.1.3. Number of registered economic organizations (in Hungarian), available at: https://www.ksh.hu/docs/hun/xstadat/xstadat_eves/i_qvd010.html (accessed 10 February 2021)

[82] MNB: Hungarian National Bank (in Hungarian), available at: https://www.mnb.hu/arfolyamok (accessed 28 June 2021)

Characteristics of Crisis Management Measures in the HR Area During the Pandemic in Hungary – Results of a Countrywide Survey of Organizations

Krisztina Dajnoki

University of Debrecen Faculty of Economics and Business, Böszörményi út 138, H-4032 Debrecen, Hungary, e-mail: dajnoki.krisztina@econ.unideb.hu

József Poór

J. Selye University Faculty of Economics and Informatics, Bratislavská cesta 3322, 945 01 Komárno, Slovalia, e-mail: poorj@ujs.sk

Ákos Jarjabka

University of Pécs Faculty of Business and Economics, Rákóczi út 80, H-7622 Pécs, Hungary, e-mail: jarjabka.akos@ktk.pte.hu

Botond Géza Kálmán

Budapest Metropolitan University, Institute of Business
Nagy Lajos király útja 1-9, 1148 Budapest, eupemq@instructor.metropolitan.hu

Zsolt Sándor Kőmüves

Hungarian University of Agriculture and Life Sciences Kaposvár Campus, Guba Sándor utca 40, H-7400 Kaposvár, Hungary, e-mail: Komuves.Zsolt.Sandor@unimate.hu

Beáta Pató Szűcs

Eötvös Loránd University Faculty of Social Sciences, Egyetem tér 1-3, H-1053 Budapest, Hungary, e-mail: szucs.beata@sek.elte.hu

Katalin Szabó

Hungarian University of Agriculture and Life Sciences Gödöllő Campus, Páter Károly utca 1, H-2100 Gödöllő, Hungary, e-mail: Szabo.Katalin@uni-mate.hu

Szilvia Szabó

Budapest Metropolitan University Institute of Business, Nagy Lajos király útja 1-9, H-1148 Budapest, Hungary, e-mail: szszabo@metropolitan.hu

Zsuzsanna Szeiner

J. Selye University Faculty of Economics and Informatics, Bratislavská cesta 3322, 945 01 Komárno, Slovakia, e-mail: szeinerzs@ujs.sk

Arnold Tóth

Budapest Business School Faculty of Finance and Accountancy, Buzogány utca 10-12, H-1149 Budapest, Hungary, e-mail: toth.arnold@uni-bge.hu

Imola Csehné Papp

Eötvös Loránd University Faculty of Education and Psychology, Kazinczy u. 23-27, H-1075 Budapest, Hungary, e-mail: papp.imola@ppk.elte.hu

András István Kun

University of Debrecen Faculty of Economics and Business Böszörményi út 138, H-4032 Debrecen, Hungary, e-mail: kun.andras.istvan@econ.unideb.hu

Abstract: The current paper is the second part of the study "Characteristics of Crisis Management Measures in the HR Area During the Pandemic in Hungary – Literature Review and Methodology". Based on two waves of a questionnaire survey conducted during the first and second waves of the COVID-19 pandemic, this article presents prevalent HRM reactions of Hungarian organisations (companies and institutions) and tests if these reactions are connected to the organisation's size (employee headcount and revenue), ownership and sector (main field of activity). The testing is conducted separately on two relatively large non-representative samples (N1 > 300, N2 > 900) employing statistical (rank-correlation,

rank-sum, and association) analyses. One of the characteristic changes the pandemic triggered in the companies and institutions is that the functions of HRM have been expanded with new activities (e.g. home office, health protection). The respondents indicated considerable new efforts, especially after the second wave, such as replacing plans, providing new employees, and addressing employees' social problems. Our research also highlighted that smaller SMEs regard it as less important to maintain and operate an active HRM function even during the pandemic. With the increase in organisational size (employee headcount and revenue) home office, the new workplace safety and occupational health measures, the hiring freeze and downsizing are substantially more frequent at larger companies as opposed to smaller ones. We are aware of the fact that the practices applied by a country of 10 million people during the pandemic have no substantial global influence; at the same time, the experiences gained and presented here can expand and diversify the methods and toolkits applied in this area.

Keywords: COVID-19; pandemic; crisis; human resource management; home office, survey

1 Introduction

The current article is the second, empirical part of the study "Characteristics of Crisis Management Measures in the HR Area During the Pandemic in Hungary – Literature Review and Methodology". The previous paper consisted of the literature review that provided a background for the research as well as the hypotheses derived from the reviewed papers and the description of the methodology of the surveys and the hypothesis testing. Therefore, we only provide a brief summary here.

The research this paper investigates the questions of what responses, solutions, and practices organisations used against the appearance of the COVID-19 pandemic and what kind of changes it has inducted, both in general and in particular within the field of Human Resource Management (HRM).

In the first part of our study, we presented a literature review describing the main effects of the first two pandemic waves on economies, organisations, and particularly on HRM contexts and practices. The empirical research hypotheses presented in the first part of the study are based on the examinations by [5] and [15]. To build a better connection between the first and the second parts of our study, we repeat the hypotheses below (see the detailed connections of the hypotheses to the literature as well as the description of the methodology in the first part of the study).

H1: The typically occurring HRM crisis management measures caused by the effects of the COVID-19 pandemic are primarily related to headcount management/workforce requirements (hiring freeze, downsizing, downsizing of temporary staff, reducing labour requirements by automation/technical solutions and the innovative handling of work hours (reduction of working hours, enabling/directing home offices, elaboration/re-planning of replacement plans).

H2: Among the responding organisations the most frequently applied HR crisis management measures are correlated with the organisation's revenue and headcount: enabling/directing home offices, new occupational health and safety measures, hiring freeze and downsizing are considerably more frequent at larger companies than at smaller ones.

The H2 hypothesis is broken down into sub-hypotheses:

H2a: Among the responding organisations the most frequently applied HR crisis management measures are correlated with the organisation's two main size indicators: revenue and headcount.

H2b: Enabling/directing home offices, new occupational health and safety measures, and reducing the risks of the pandemic HR crisis management measures are correlated with the organisation's two main size indicators: revenue and headcount.

H2c: The "No task" attitude is typical mainly at organisations that are smaller according to revenue and headcount.

We divided the reporting of the result into two sub-sections by the two hypotheses. In both parts we follow the methodology described in the paper "Characteristics of Crisis Management Measures in the HR Area During the Pandemic in Hungary – Literature Review and Methodology": frequency and rank-correlation and associational analyses, and rank-sum tests are conducted.

The data were collected in two phases of a questionnaire survey during the first two waves of the pandemic (between 12 June and 31 July 2020 and between 01 August and 15 November 2020), resulting in 508 and 1,041 responses, respectively. Since the samples during the two survey phases do not match, the results are not directly comparable; thus, we analyze them in the present study separately. The demographics of the samples are described in the first part of our article. The same questionnaire was used during both survey phases (in original language: http://limesurvey.szie.hu/index.php/44678?lang=hu; the authors will provide the English translation electronically on request).

The methods of analysis are kept simple, but are adequate for testing our hypotheses (for the detailed description, see the Material and methodology section of the first part of our article).

Conclusions and the discussion of the study's limitations close the article.

2 Results from Two Questionnaire Survey Waves

In the current section, we are presenting the tests of the two main (and three sub-) hypotheses in two consecutive sub-sections.

2.1 Examination of the First Hypothesis

In the first phase of our research, – which comprises the examination of the first wave of the pandemic – we investigated what HR crisis management measures were introduced by the respondent organisations caused by the appearance of the coronavirus pandemic. Figure 1 describes the typical crisis management measures identified by the research results.

An overwhelming majority of the respondent organisations (86%) concluded at the first wave of the pandemic that some measures were necessary for the HR area, while only 13.8% indicated that they characteristically did not have any HR tasks at the beginning of the crisis situation.

New occupational health and work safety measures were indicated in the highest proportion (82%) as a task that appeared to some extent, of which a considerable percentage of respondents (42%) indicated that it was highly characteristic. Enabling/directing home offices was present in a similarly high percentage (81%) in the practice of the organisations, in such a way that over half of them (53%) selected the highly characteristic category.

In the sample elaboration/re-planning of replacement plans, and addressing employees' social problems appeared the third and fourth place as HR interventions with similar percentages (71-69%).

The next two places among somewhat typical solutions are supporting personal development (57%) and hiring freeze (55%). At the former, the highly typical marking is 14%, while at the latter it is higher, 30%.

It can be established that among the studied 19 HR measures, respondents marked a dozen as non-existent in their practice. There are (in the least typical order): revising the equal opportunities strategic plan (82%), pay reduction (78%), downsizing of temporary staff (76%), reducing labour requirements by training, development (75%), reducing labour requirements by automation/technical solution (73%), reducing fringe benefits (73%), revising the performance appraisal system (70%), downsizing (65%), revising the incentive scheme (65%), pay freeze (62%), reduction of working hours (57%) reducing the risks of the pandemic through training (56%).

In the survey of the second wave of the pandemic, we also examined what HR crisis management measures were introduced by the respondent organisations caused by the appearance of the coronavirus pandemic.

Figure 2 summarises the data of the second wave and presents what HR crisis management measures characterised the respondent organisations, and to what proportion of the organisations the measures extended to.

A substantial percentage of organisations (69.1%) considered that HR had tasks related to the effects of the pandemic. I.e. less than one-third of them did not see a reason for intervention by the HR area to a low, medium, or high extent. Of these

8.3% of respondents very strongly did not see HR tasks appear in relation to the pandemic.

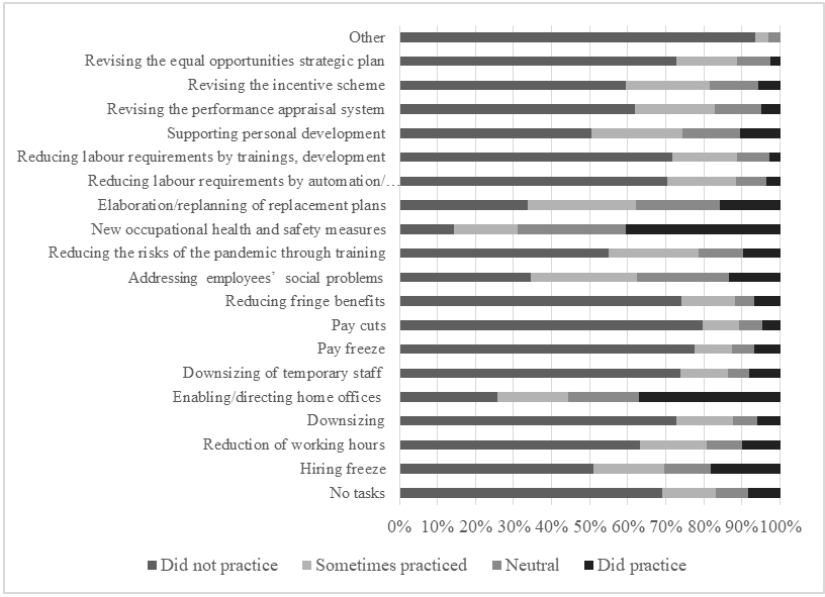


Figure 1
Typical crisis management HR measures in the first phase

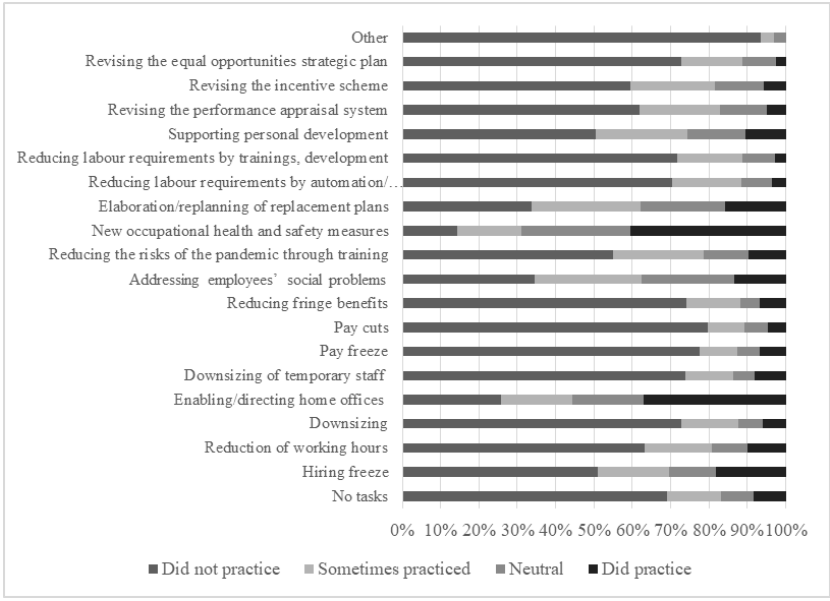


Figure 2
Typical crisis management HR measures during the second wave

The measures marked by most respondents as characteristic at least to a low extent: "New occupational health and safety measures" (86%), "Enabling/directing home offices" (74%), "Elaboration/re-planning of replacement plans" (66%) and "Addressing employees' social problems" (65%). But for nearly 50% of organisations (45-50%) "Supporting personal development", "Hiring freeze" and "Reducing the risks of the pandemic through training" was also typical. At 30-41% of organisations "Revising the incentive scheme", "Revising the performance appraisal system" (the score of these being close to each other is a good sign and may suggest an integrated approach to these two areas), and "Reduction of working hours" to some extent. Less than 30% of respondents indicated "Reducing labour requirements by automation/technical solution", "Reducing labour requirements by training, development", "Downsizing", "Revising the equal opportunities strategic plan", "Downsizing of temporary staff" and "Reducing fringe benefits". The two last places with under 25% were "Pay freeze" and "Pay reduction". These percentages show that (at least based on the answers of the respondents) proactive and forward-looking solutions, those seeing the remedy in development, were more frequent. It may be separately highlighted that from the ranking the effort is evident to provide production capacity by physical or human capital investment (technological development, training). These are forward-looking solutions even without the pandemic situation, but the current situation may have given them momentum.

If we only consider the highly typical markings, then in the second wave of the pandemic the most frequent HR measures were „New occupational health and safety measures" (28%), "Addressing employees' social problems" (24%), "Elaboration/re-planning of replacement plans" (22%). Furthermore, marked 10-20% were "Enabling/directing home offices", "Supporting personal development", "Revising the incentive scheme", "Hiring freeze", "Revising the performance appraisal system" and "Reducing the risks of the pandemic through training". At this level of importance, other HR solutions were under 10%.

2.2 Examination of the Second Hypothesis

According to the description in the methodology chapter, we first conducted rank correlation studies, specifically on each sample separately and in chronological order.

The rank correlation with the headcount variable is shown in diminishing order of strength in Table 1, omitting the non-significant variable pairs.

Table 1

Rank correlation between HR measures and headcount (first sample)

Measures	***N***	**Spearman's rho**
New occupational health and safety measures	363	0.419***
Elaboration/re-planning of replacement plans	360	0.304***

Downsizing of temporary staff	358	0.287***
No tasks	*339*	*–0.250***
Enabling/directing home offices	363	0.247***
Addressing employees' social problems	362	0.223***
Hiring freeze	360	0.190***
Reducing labour requirements by automation/ technical solution	359	0.175**
Reducing labour requirements by training, development	357	0.141**
Reducing the risks of the pandemic through training	360	0.137**

* $p < 0.050$, ** $p < 0.010$, *** $p < 0.001$, rank correlation cases with over 0.200 absolute value marked with a grey background, negative values with italics

The rank correlation with the revenue variable is shown in diminishing order of strength in Table 2, omitting the non-significant variable pairs.

Table 2

Rank correlation between HR measures and revenue (first sample)

Measures	***N***	**Spearman's rho**
New occupational health and safety measures	358	0.414***
Elaboration/re-planning of replacement plans	355	0.277***
Downsizing of temporary staff	358	0.267***
No tasks	353	0.266***
Enabling/directing home offices	*334*	*–0.233***
Addressing employees' social problems	357	0.182**
Hiring freeze	355	0.161**
Reducing labour requirements by automation/ technical solution	355	0.160**
Reducing labour requirements by training, development	354	0.120*
Reducing the risks of the pandemic through training	352	0.115*

* $p < 0.050$, ** $p < 0.010$, *** $p < 0.001$, rank correlation cases with over 0.200 absolute value marked with a grey background, negative values with italics

The rank correlation with the headcount variable is shown in diminishing order of strength in Table 3, omitting the non-significant variable pairs.

Table 3

Rank correlation between HR measures and headcount (second sample)

Measures	***N***	**Spearman's rho**
Enabling/directing home offices	979	0.323***
New occupational health and safety measures	974	0.269***
Reducing the risks of the pandemic through training	973	0.253***
Elaboration/re-planning of replacement plans	972	0.242***
Reducing labour requirements by training, development	967	0.201***
Addressing employees' social problems	975	0.187***

Downsizing of temporary staff	974	0.173***
Hiring freeze	973	0.152***
Reducing labour requirements by automation/ technical solution	968	0.151***
Revising the equal opportunities strategic plan	968	0.138***
Revising the performance appraisal system	969	0.122***
Supporting personal development	971	0.106**
No tasks	*926*	*–0.096***
Revising the incentive scheme	970	0.085**
Pay freeze	977	0.081*
Downsizing	977	0.079*

* $p < 0.050$, ** $p < 0.010$, *** $p < 0.001$, rank correlation cases with over 0.200 absolute value marked with a grey background, negative values with italics

The rank correlation with the revenue variable is shown in diminishing order of strength in Table 4, omitting the non-significant variable pairs.

Table 4

Rank correlation between HR measures and revenue (second sample)

Measures	***N***	**Spearman's rho**
Enabling/directing home offices	957	0.328***
New occupational health and safety measures	952	0.313***
Elaboration/re-planning of replacement plans	950	0.203***
Addressing employees' social problems	953	0.195***
Reducing the risks of the pandemic through training	951	0.186***
Downsizing of temporary staff	952	0.149***
Reducing labour requirements by training, development	945	0.145***
Hiring freeze	951	0.139***
Reducing labour requirements by automation/ technical solution	946	0.106**
Revising the performance appraisal system	947	0.098**
Supporting personal development	949	0.096**
No tasks	*906*	*–0.083**
Pay cuts	*955*	*–0.082**
Pay freeze	955	0.079*
Revising the incentive scheme	948	0.068*

* $p < 0.050$, ** $p < 0.010$, *** $p < 0.001$, rank correlation cases with over 0.200 absolute value marked with a grey background, negative values with italics

Chi-square (cross-table) analyses for the H2c hypothesis

At the actual and random distribution according to the headcount categories, the expected frequencies for the first sample are shown in Table 5.

Table 5

Considering the "No tasks" measure characteristic according to headcount categories (first sample)

How characteristic is the No task opinion?			Headcount categories (person)							Total
			0	1-9	10-49	50-250	251-500	501-2000	over 2000	
No task at all	No	Actual	8	44	42	45	27	28	14	208
		Expected	11.7	58.9	41.7	38.0	20.9	22.7	14.1	208.0
	Yes	Actual	11	52	26	17	7	9	9	131
		Expected	7.3	37.1	26.3	24.0	13.1	14.3	8.9	131.0
Medium level no task	No	Actual	11	63	52	55	30	33	19	263
		Expected	14.7	74.5	52.8	48.1	26.4	28.7	17.8	263.0
	Yes	Actual	8	33	16	7	4	4	4	76
		Expected	4.3	21.5	15.2	13.9	7.6	8.3	5.2	76.0
Very much no task	No	Actual	11	77	59	58	32	33	21	291
		Expected	16.3	82.4	58.4	53.2	29.2	31.8	19.7	291.0
	Yes	Actual	8	19	9	4	2	4	2	48
		Expected	2.7	13.6	9.6	8.8	4.8	5.2	3.3	48.0

According to the Table's data, in the two lowest headcount categories (0 and 1-9) the received frequencies are always below the expected value, meaning that the smallest organisations, in comparison to random distribution, answered more frequently than some extent they agree that they have No HR tasks. In the case of No task at all the value of Cramér's V is 0.266 ($p = 0.001$), thus here the correlation is significant. In the case of Medium level no task 1, at the Very much no task 3 expected cell values did not reach 5, thus here the test could not be performed. To remedy this, we created a dummy variable based on the headcount, with the value 1 if the organisation's headcount does not exceed 9, and with the value 0 if it is higher. By the application of this variable on a 2x2 matrix the test could be performed in the case of Medium level no task ($V = 0.227$, $p < 0.001$) and the Very much no task ($V = 0.192$, $p < 0.001$), with a significant result in both cases (to avoid data duplication we did not publish these Tables, since they are clear from the above Table).

At the actual and random distribution according to the revenue categories, the expected frequencies are shown in Table 6.

Table 6

Considering the "No tasks" measure characteristic according to revenue categories (first sample)

How characteristic is the No task opinion?	Annual budget/revenue in 2019							Total
	under 50 million HUF	51 - 100 million HUF	101 - 500 million	501 million HUF - 2.5	2.51 - 25 billion	25.1 - 120 billion	over 100 billion HUF	

					HUF	billion HUF	HUF	HUF		
No task at all	**No**	Actual	40	16	32	29	60	15	12	204
		Expected	51.9	19.5	34.2	26.3	46.4	14.7	11.0	204.0
	Yes	Actual	45	16	24	14	16	9	6	130
		Expected	33.1	12.5	21.8	16.7	29.6	9.3	7.0	130.0
Medium level no task	**No**	Actual	53	21	45	34	70	21	14	258
		Expected	65.7	24.7	43.3	33.2	58.7	18.5	13.9	258.0
	Yes	Actual	32	11	11	9	6	3	4	76
		Expected	19.3	7.3	12.7	9.8	17.3	5.5	4.1	76.0
Very much no task	**No**	Actual	63	27	50	36	73	23	14	286
		Expected	72.8	27.4	48.0	36.8	65.1	20.6	15.4	286.0
	Yes	Actual	22	5	6	7	3	1	4	48
		Expected	12.2	4.6	8.0	6.2	10.9	3.4	2.6	48.0
Total		Actual	85	32	56	43	76	24	18	334
		Expected	85.0	32.0	56.0	43.0	76.0	24.0	18.0	334.0

* 1 EUR = 351.1 HUF [17]

The lessons of distribution according to the revenue categories are similar to the study conducted based on headcount. In the case of the at least low extent agreement those with revenue under HUF 501 million, at the other two dummy variables, those with revenue under HUF 101 million agreed more frequently than what would be expected without correlation. In the first case, the precondition of Cramér's V calculation is met, and according to this, the correlation was significant ($V = 0.246$, $p = 0.003$). Similarly to the solution applied in the case of headcount, we made a 2x2 matrix here as well. The values of the new revenue-based dummy were 0, if the indicated revenue was under HUF 100 million, and 1 if it was over that amount. The test runs in this way had a significant result in both cases (in the case of medium-level agreement: $V = 0.245$, $p < 0.001$; at a high-level: $V = 0.182$, $p = 0.001$).

Thus, in the first sample, we found evidence that H2c cannot be disregarded.

At the actual and random distribution according to the headcount categories, the expected frequencies are shown in Table 7.

Table 7

Considering the "No tasks" measure characteristic according to headcount categories (second sample)

How characteristic is the No task opinion?			**Headcount categories (person)**							**Total**
			0	**1-9**	**10-49**	**50-250**	**251-500**	**501-2000**	**over 2000**	
No task at all	**No**	Actual	19	94	135	151	58	95	89	641
		Expected	18.0	112.1	139.1	139.1	59.5	87.2	85.8	641.0
	Yes	Actual	7	68	66	50	28	31	35	285
		Expected	8.0	49.9	61.9	61.9	26.5	38.8	38.2	285.0
Medium level no task	**No**	Actual	23	120	165	176	72	112	102	770
		Expected	21.6	134.7	167.1	167.1	71.5	104.8	103.1	770.0
	Yes	Actual	3	42	36	25	14	14	22	156
		Expected	4.4	27.3	33.9	33.9	14.5	21.2	20.9	156.0
Very much no task	**No**	Actual	24	138	182	191	80	123	112	850
		Expected	23.9	148.7	184.5	184.5	78.9	115.7	113.8	850.0
	Yes	Actual	2	24	19	10	6	3	12	76
		Expected	2.1	13.3	16.5	16.5	7.1	10.3	10.2	76.0
Total		Actual	26	162	201	201	86	126	124	926
		Expected	26.0	162.0	201.0	201.0	86.0	126.0	124.0	926.0

In comparison with the first sample, it is a conspicuous difference that the organisations employing 0 people agreed the most frequently that HR tasks are unnecessary than what would be expected in the case of random distribution, moreover even less frequently. However, organisations employing 1-9 people showed characteristics similar to the earlier, and even organisations with less than 50 people seemed to „join" them. The Cramér's V calculation was only possible in the case of the first variable (V = 0.139, p = 0.013), which was significant. We transformed the headcount variable into a dummy again, but now (in light of the new tasks) we drew the line at 50 people rather than 10. In this way, the correlation was significant in the case of the Medium level no task (V = 0.090, p = 0.006) as well as the Very much no task (V = 0.104, p = 0.002). Thus, in summary, we are witnessing a phenomenon similar to the data collection during the first wave, with the difference that the behavior of organisations with no employees did not match the previous and the headcount limit was pushed higher.

At the actual and random distribution according to the revenue categories, the expected frequencies are shown in Table 8.

Table 8
Considering the "No tasks" measure characteristic according to revenue categories (second sample)

How characteristic is the No task opinion?			Annual budget/revenue in 2019: under 50 million HUF	51 - 100 million HUF	101 - 500 million HUF	501 million HUF - 2.5 billion HUF	2.51 - 25 billion HUF	25.1 - 120 billion HUF	over 100 billion HUF	Total
No task at all	**No**	Actual	129	59	97	124	101	51	66	627
		Expected	139.8	63.0	105.9	114.2	92.7	47.8	63.7	627.0
	Yes	Actual	73	32	56	41	33	18	26	279
		Expected	62.2	28.0	47.1	50.8	41.3	21.2	28.3	279.0
Medium level no task	**No**	Actual	159	77	132	142	114	55	76	755
		Expected	168.3	75.8	127.5	137.5	111.7	57.5	76.7	755.0
	Yes	Actual	43	14	21	23	20	14	16	151
		Expected	33.7	15.2	25.5	27.5	22.3	11.5	15.3	151.0
Very much no task	**No**	Actual	179	87	144	152	123	64	84	833
		Expected	185.7	83.7	140.7	151.7	123.2	63.4	84.6	833.0
	Yes	Actual	23	4	9	13	11	5	8	73
		Expected	16.3	7.3	12.3	13.3	10.8	5.6	7.4	73.0
Total		Actual	202	91	153	165	134	69	92	906
		Expected	202.0	91.0	153.0	165.0	134.0	69.0	92.0	906.0

* 1 EUR = 351.1 HUF [17]

The result of the chi-square (cross-table) analyses performed on revenue categories showed marked differences between the two waves as well. In the second, it was only true for the lowest revenue category in the case of all three agreement variables that more respondents marked the proper level of the No task option than larger organisations. But in the case of No task, at all even the organisations under HUF 500 million agreed more with no action than the larger ones. The Cramér's V calculation was possible in the case of all three variable pairs, but it did not show a significant correlation in any of them. After transforming the revenue variables into dummies we were able to show significant differences according to revenue categories. In the case of No task at all the limit was HUF 500 million ($V = 0.113$, $p = 0.001$), HUF 50 million with Medium level no task ($V = 0.066$, $p = 0.046$), and Very much no task ($V = 0.066$, $p = 0.049$).

Conclusions

According to our first hypothesis (H1), the crisis management measures that typically occur in the HR sector due to the effects of the pandemic are primarily related to headcount management/workforce requirements. In the first phase of the research, respondents indicated the New occupational and health safety measures as the highest percentage, and the Enabling / directing home office was included in the practice of the organisations in a practically similar proportion. These measures were clearly necessary, as the cause of the crisis was a health problem attacking human resources, the coronavirus. Addressing employees 'social problems came in third, which in turn clearly indicates a paradigm shift. Downsizing used almost as a template in the management of previous crises was not among the top ten in the frequency ranking of the measures examined. Instead, Pay freeze, Hiring freeze, and Reducing working hours were chosen by the organisations surveyed. This is related both to addressing social problems and to the fact that companies have realised the danger of loss of knowledge due to Downsizing. So headcount management was really important, but HR has already done this based on a new paradigm. The growing importance of human capital measures is, therefore, in line with international experience [13].

In the crisis caused by the pandemic, many organisations focused on the safety, health, and well-being of employees [7] [22]. They tried to respond to the challenges with thoughtful planning and the development of employee's skills and competencies [12] [18]. The frequency of remote work has increased significantly worldwide [14]. This was also the most common reaction in our survey. Overall, the frequency of this was significantly higher in companies with more employees, especially in those with foreign ownership, which is also consistent with the literature [3]. However, in the first phase of the epidemic.it was precisely the large companies that introduced remote work less often [2], which is supported by the results of our study.

According to our second research hypothesis (H2), the most commonly used HR crisis management measures in the sample correlate with the organisation's revenue

and headcount: Enabling/directing home office [16]. New occupational and health safety measures [21]. Hiring freeze and Downsizing are significantly more common in larger firms than in smaller ones [9]. The rank correlation analysis, supplemented by a cross-tabulation analysis for the third sub-hypothesis, confirmed that the larger the relative headcount (H2a) and revenue (H2b) of an organisation, the more prone it is to practically all of the measures listed. The word "practically" is needed because Downsizing showed a significant correlation only for temporary workers.

The very weak correlation found between the "No" tasks attitude and the headcount category, and the revenue category shows that this attitude is more typical for smaller organisations. Cross-tabulation analyses have confirmed this, but also indicate that in this respect the limit between small and large organisations may have been different at each phase of the pandemic. This is also consistent with the results of international studies [1].

According to our results, nearly 10% of companies did not respond to the crisis at all. International data, on the other hand, show that SMEs, which are more common in our sample, responded more quickly and proactively compared to large companies [6], and their employees were more satisfied with the measures [10].

Among the typical HR measures, wage reductions were more typical of small companies, downsizing more for large ones [1]. Both wage cuts and downsizing were more typical of foreign-owned companies than domestic companies. It was observed that both previously indicated measures were proportional.

As the correlation in our results is weak, further research would be needed to substantiate this statement.

Limits and Future Plans

In the first and second phases of our research (i.e. the two waves of a questionnaire survey), which roughly coincided with the first and second waves of the pandemic, we reviewed the activities of nearly one and a half thousand Hungarian companies and institutions. We are fully aware that our sample is not representative. However, it is relatively large and heterogenous, thus the organisations participating in it represents different parts of the Hungarian economy.

According to [4] Budhwar and Cumming (2020), the COVID-19 crisis brought attention to the importance of an international perspective. In their view, the pandemic reminded us of the interconnectedness of the entire world. Therefore, studies from certain regions – such as the Hungarian example in this article – may provide an important local perspective on the pandemic's organisational management and may help us find global solutions. Our future plans include expanding the study in an international dimension: We started our research using an online survey in other eastern and central European countries such as Austria, Slovakia, Bosnia, and Romania. It can be stated that organisational responses to a pandemic must also consider national-organisational cultural differences.

As a conclusion of our article, let us mention the opinion regarding COVID of [8] Diamond (2020:13) the world-famous researcher of the effects crises and pandemics, 'the real threat to our civilisation is not posed by COVID, but rather by climate change, exhausting resources, and global inequalities".

References

[1] Achou, B. D. Boisclair, P. d'Astous, R. Fonseca, F. Glenzer, P.-C. Michaud: Early Impact of the COVID-19 Pandemic on Household Finances in Quebec. Canadian Public Policy, 46(s3), 217-235, https://doi.org/10.3138/cpp.2020-087

[2] Bartik, A. W., Cullen, Z. B., Glaeser, E. L., Luca, M.,Stanton, C. T. (2020) *What Jobs are Being Done at Home During the Covid-19 Crisis? Evidence from Firm-Level Surveys* (p. 26)

[3] Borino, F., Carlson, E., Rollo, V., Solleder, O. (2021) International firms and COVID-19: Evidence from a global survey. *Covid Economics*, *75*(7), 30-60, https://www.researchgate.net/profile/Michiru-Kaneda-2/publication/351942767_

[4] Budhwar, P., Cumming, D: New Directions in Management Research and Communication: Lessons from the COVID-19 Pandemic, British Journal of Management, 2020, Vol. 31, No. 3, pp. 441-443

[5] Collings, D. D. Mcmakin, J., Nyberg, J. A., Wright, M. P. (2021): Strategic Human Resource Management andCOVID-19: Emerging Challenges and ResearchOpportunities., *Journal of Management Studies*, 58(5), 2-18, DOI: 10.1111/joms.12695

[6] Costa, S., Santis, S. D., Monducci, R. (2022) Reacting to the COVID-19 crisis: State, strategies and perspectives of Italian firms. *Rivista Di Statistica Ufficiale/Review of Official Statistics*, *2022*(1) 73-107

[7] Dennerlein, J. T., Burke, L., Sabbath, E. L., Williams, J. A. R., Peters, S. E., Wallace, L., Karapanos, M., Sorensen, G. (2020) An Integrative Total Worker Health Framework for Keeping Workers Safe and Healthy During the COVID-19 Pandemic. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *62*(5) 689-696, https://doi.org/10.1177/0018720820932699

[8] Diamond, J. (2020), Turning Points from Nations in Crisis (in Hungarian), Akkord Publishing House, Budapest, Hungary. 13. The English version published in 2019, Diamond, J. (2019) Upheaval. New York: Little Browen Co

[9] European Commission (2021): Telework in the EU before and after the COVID-19: Where we were, where we head to (No. JCR120945; p. 8). European Commission, 2020, https://ec.europa.eu/jrc/sites/default/files/jrc120945_policy_brief_-_covid_and_telework_final.pdf

[10] Frutos-Bencze, D., Sokolova, M., Zubr, V., Mohelska, H. (2022) Job Satisfaction During Covid-19: Industry 5.0 as a Driver of Sustainable Development and Gender Equality. *Technological and Economic Development of Economy*, *28*(5), 1527-1544, https://doi.org/10.3846/tede.2022.17680

[11] Hamouche S (2021) Human resource management and the COVID-19 crisis: implications, challenges, opportunities, and future organizational directions. *Journal of Management & Organization* 1-16, https://doi.org/10.1017/jmo.2021.15

[12] Hamza Shuja, K., Aqeel, M., Foundation University Islamabad, Rawalpindi Campus, Islamabad, Pakistan, Jaffar, A., & Ahmed, A. (2020) Covid-19 Pandemic and Impending Global Mental Health Implications. *Psychiatria Danubina*, *32*(1), 32-35, https://doi.org/10.24869/psyd.2020.32

[13] ILO (2020) A safe and healthy return to work during the COVID-19 pandemic available at: https://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---safework/documents/briefingnote/wcms_745549.pdf (accessed 7 July 2021)

[14] Koirala, J., Acharya, S. (2020) Dimensions of Human Resource Management Evolved with the Outbreak of COVID-19. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3584092

[15] Lazányi, K. (2018): A bizalom hatása a munkaerő-piaci helyzetre. KÖZÉP-EURÓPAI KÖZLEMÉNYEK, 11(3), 128-138

[16] Lewis, N.: HR Managers Rethink Their Role During the Coronavirus Pandemic" available at: https://www.shrm.org/hr-today/news/hr-news/pages/hr-managers-rethink-their-work-coronavirus-pandemic.aspx (accessed 1 June 2021) MNB: Hungarian National Bank (in Hungarian)

[17] MNB (2021): Hungarian National Bank (in Hungarian) available at: https://www.mnb.hu/arfolyamok (accessed 28 June 2021)

[18] Risley, C. (2020) Maintaining Performance and Employee Engagement During the COVID-19 Pandemic. *Journal of Library Administration*, *60*(6), 653-659, https://doi.org/10.1080/01930826.2020.1773716

[19] Shinozaki, S., Rao, L. N. (2021) COVID-19 Impact on Micro, Small, and Medium-Sized Enterprises under the Lockdown: Evidence from a Rapid Survey in the Philippines. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3807080

[20] Soomro, M. A., Ali, M. H., Tan, K. H., Kumar, A., Sinniah, S. (2022) Workforce resilience in the post-COVID-19 era: Differences based on manufacturing–service orientation and firm size. *Production Planning & Control*, 1-13, https://doi.org/10.1080/09537287.2022.2106446

[21] Weidemeyer, F.: COVID-19 crisis management: ten better questions to ask available at: https://www.ey.com/en_gl/covid-19/covid-19-crisis-management-essential-ten-better-questions-to-ask (accessed 1 June 2021)

[22] Wu, L.-K., Su, W.-H., Hsiao, S.-H., Hou, M.-F. (2020) Preparedness for the next emerging infectious disease outbreak by implementing strategic human resource management. *Journal of the Chinese Medical Association*, *83*(10), 973-974, https://doi.org/10.1097/JCMA.0000000000000423

Water 4.0 in Hungary: Prospects and Cybersecurity Concerns

Tamás Szádeczky

Department of Instrumentation and Automation, Institute of Electronic and Communication Systems, Kandó Kálmán Faculty of Electrical Engineering, Óbuda University, Bécsi út 96/b, 1034 Budapest, Hungary
szadeczky.tamas@kvk.uni-obuda.hu

Budapest University of Technology and Economics, Faculty of Economic and Social Sciences, Department of Management and Business Economics, Műegyetem rkp. 3, 1111 Budapest, Hungary
szadeczky.tamas@gtk.bme.hu

Abstract: This paper discusses the increasing significance of smart water management, within the context of the fourth industrial revolution and the associated cybersecurity risks, particularly in Hungary and Central Europe. By examining the current state of smart water management and analyzing the various cybersecurity threats, this study seeks to raise awareness around the need for enhanced security measures, in this critical sector. The research methodology is primarily based on a through literature review and secondarily, on related data analysis. The paper identifies several cybersecurity challenges and potential solutions for smart water management and finally suggests future research directions, to ensure the safe and sustainable development of this critical infrastructure.

Keywords: Control system security; Critical infrastructure protection; Cybersecurity; Digital transformation; Water utility

1 Introduction

The role of IT in everyday life in the 21st Century is undeniable. Technology makes our lives more convenient and efficient, with less repetitive work. Machines can effectively automate processes. This technological shift leads to higher dependence on IT at the level of the individual, the organization and society. At the level of the individual, the vast majority of tasks now require the use of some kind of a computer. We can do our work from home, connected to the corporate network via a VPN or perhaps, on our tablet, while sitting on the plane. We post our queries to Google search and e-mail our business partners, preferably electronically signed and encrypted. We reach our remote loved ones via teleconferencing or chat

applications. We used these facilities in our daily work, during the peak virus season extensively.

When we think of an organization's dependence on IT, we can associate it with virtually any business record. For example, customer records are now kept in customer relationship management (CRM) systems. Accounting records are also stored on computers. While single-entry accounting, cash book, can still be done on paper, for double-entry accounting this now seems completely anachronistic. Above medium size, it is common for organizations to introduce an integrated ERP system, such as SAP or Microsoft Dynamics, which is used, not only for tracking costs, but also in the production process.

1.1 Industrial Revolutions

The term 'Industrial Revolution' is the best way to describe the automation of production and manufacturing. Such revolutions represented major advances in technology and significantly impacted both economy and society.

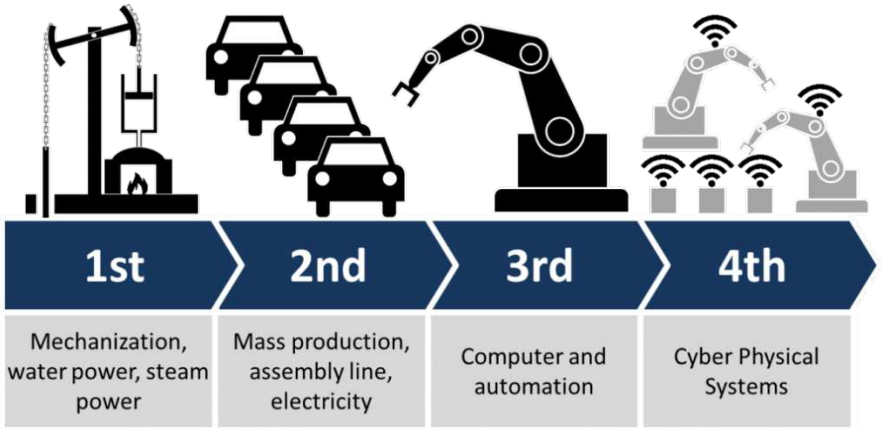


Figure 1
Industrial revolutions [1]

We count the first industrial revolution from the advent of the steam engine. This was strongly linked to the bourgeois revolution and led to the decline of the feudal system throughout Europe. At the technological level, it was characterized by the mechanization of the textile industry, steam-powered ships, steam railways, and the use of steam engines in manufacturing [2]. The second industrial revolution saw the emergence of mass production. One example of this was the Ford car company's mass production of motor vehicles. However, with the development of the iron and steel industry, the use of electricity, the invention of the combustion engine, the chemical industry, agricultural development, and the military industry also underwent significant technical developments. Economically, it was characterized

by the emergence of monopolies and, socially, by the emergence of a middle class and the differentiation of the working class. The third industrial revolution was the advent of computer technology and thus the automation of processes. This involved autonomous regulation and control and some form of centralized coordination, i.e., the industrial control systems (ICS). One of ICS's basic components is the programmable logic controller (PLC), which is in direct contact with sensors and actuators integrated into the process being controlled. These controllers can be interconnected over a suitable network, formerly RS-485, today Industrial Ethernet. Measurement data can be collected, for example, by a supervisory control and data acquisition (SCADA) system, or distributed control can be implemented by distributed control systems (DCS).

The fourth industrial revolution, also referenced as Industry 4.0, or Industrial Internet of Things (IIoT), is currently underway. It is marked by the convergence of digital, physical, and biological systems, as well as the emergence of new technologies such as the Internet of Things (IoT), Artificial Intelligence (AI), robotics, and big data analytics [3]. This industrial revolution is based on these technologies but uses different, modern information technology principles and methods and is a data-driven and networked approach to manufacturing [4]. The IT tools of the 21st Century can collect and analyze previously unimaginable amounts of data, which is what data science or data science is all about. In addition, the massive use of the web developed in the 2000s – making the internet a ubiquitous and quasi-ubiquitous tool for all our lives – is also spilling over into manufacturing and process management. This will make manufacturing even more efficient. At any given moment we can optimize the process in real-time by analyzing millions of data sources, making more efficient use of human resources and becoming more effective at getting the job done. Whether this really is an industrial revolution, with economic and social impact, or just a passing fad that exists only on salespersons' desks can only be determined from a historical perspective. However, what is certain is that the use of IT or, more pejoratively, dependence on IT, is a seemingly unstoppable process.

However, some papers mention the existence of the fifth industrial revolution or Industry 5.0, which is a controversial topic and is not examined in this paper.

Smart water management, as a specialized application of Industry 4.0, involves using advanced digital technologies and data-driven approaches to optimize the planning, operation, and maintenance of water resources and infrastructure and improve the efficiency, sustainability, and resilience of water services and ecosystems. In recent years, smart water management has gained increasing attention and investment from more governments (mostly in well-developed states), utilities, industries, and some research institutions worldwide, driven by the growing challenges of water scarcity, pollution, climate change, and aging infrastructure, as well as the increasing expectations of consumers, regulators, and stakeholders for better quality, affordability, and accountability in water management [5].

However, the emergent reliance on those digital technologies, automation, and connectivity in smart water management also exposes the water sector to new cybersecurity threats, vulnerabilities, and risks, well known in IT security, which can have significant consequences for the safety, security, and reliability of critical water infrastructure and services, as well as for public health, the environment, and the economy. The increasing interdependence and complexity of water systems, combined with the rapid pace of technological innovation, deployment and the evolving nature of cyber threats, make it essential for researchers, practitioners, and policymakers to understand better, assess, and manage the emerging cybersecurity challenges and opportunities in intelligent water management.

1.2 Objectives, Research and Methodology

The paper aims to contribute to the growing body of knowledge and debate on cybersecurity and smart water management by examining the current state of smart water management in Hungary and Central Europe and by analyzing various cybersecurity threats, vulnerabilities, and risks in this critical sector. The paper also proposes a set of research questions and recommendations for future studies and actions to enhance the cybersecurity and resilience of smart water management in the region.

The research questions addressed in this study include:

- What are the drivers for digitalization in the water sector?
- What are the main cybersecurity vulnerabilities in the digitalization of water management in Hungary?
- What protection solutions are currently in place to address these vulnerabilities?
- What future research directions can be identified to strengthen the security of critical water infrastructure in Hungary?

The research methods involve a comprehensive literature review of academic articles, books, reports, and other publications related to the digitalization of water management, cybersecurity issues, and the Hungarian context.

The main objective of this paper is to investigate the cybersecurity issues arising from the digitalization of water management in Hungary and to provide insights and recommendations that may apply to other Central European countries facing similar challenges.

The remainder of the paper is organized as follows: Section 2 provides an overview of the digital transformation of water management, focusing on the key concepts and technologies associated with Water 4.0. Section 3 discusses the main cybersecurity issues and concerns in smart water management. Section 4 is about the known incidents in the field and possible countermeasures, based on the

literature research and analysis. In Section 4, the implications and recommendations are discussed, with a particular emphasis on the Hungarian context and the findings of the research. Finally, Section 5 offers conclusions and recommendations for future research.

2 Smart Water Management

When someone thinks of industry 4.0, they might first associate it with a Tesla factory or some space technology center. However, the fact is that the tools needed to make it happen become available to a wide range of people in the economy and society as their price has fallen dramatically. It is enough to think that millions of users have installed smart home solutions. For an investment of a few thousand dollars, we can automate and voice-control our family home with a dozen sensors and a cloud-based personal assistant. Furthermore, based on the concept of the Smart City, public services in the municipality can be optimized and used more efficiently. One such solution could be the introduction of smart parking, where sensors embedded in the curbside pass data on to the drivers' navigation system and allow them to see the available parking spaces closest to their destination via a mobile app and thus, minimize the time spent searching for a parking space.

2.1 Digitalization Drivers of Water Management

In recent years, more countries are heading towards smart water management, which countries face various water-related challenges and opportunities, such as water scarcity, pollution, climate change, aging infrastructure, and increasing demand for high-quality, affordable, and sustainable water services. The adoption and implementation of smart water management in the Central European region have been influenced and shaped by various factors, including national and international policies, regulations, and initiatives, as well as technological, organizational, and market developments and trends [5].

For water utilities, this may sound utopistic for those who are aware of the current technical and funding situation in Central Europe. Even according to the government, the waterworks infrastructure in Hungary is in a highly depreciated state. According to a presentation by the Deputy Secretary of State for Sustainable Development at the Ministry of Innovation and Technology, Hungary [6], 30% of domestic drinking water pipelines are classified as at-risk and 56% as predominantly at-risk, while 3% of wastewater pipelines are classified as at-risk and 4% as predominantly at-risk. During the research, it was not possible to collect precise information on the degree of automation. In a case study analysis, which was done in the example of the Hungarian city of Baja, no central process control was available at the water supply of a 40000 inhabitant supply area, around the city of Baja. The government

modernized this town's infrastructure in 2012 from an EU fund, which cost 15 million euros, but still, no remote monitoring and intervention capability was built. The only way of monitoring, is that the local control equipment (PLC) sends fault reports to the operators via text messages in all water utilities. If necessary, troubleshooting is carried out with the involvement of the electrician-engineer, technologist or branch manager. Fault rectification is reported to the dispatching service, where the matter is dealt with without any process automation. Due to the deficient level of automation, in fact, the solutions are stuck at the technology level of the 90s and therefore, a significant part of the operators, also think that the issue of Water 4.0 is very far from Hungarian practice. However, incorporating the available extremely low-cost technology makes it possible to achieve much more efficient operating parameters with a low investment volume. Depending on the current IT and plant management infrastructure of a given company, this technology, or elements of it, may become part of daily practice. We should certainly expect technological developments in this direction in the coming years [7].

Despite the progress and potential of smart water management, there are still various barriers and challenges to its broader adoption and impact, such as the lack of awareness, knowledge, and skills about smart water technologies and benefits among decision-makers, stakeholders, and users; the high costs, risks, and uncertainties of investing in smart water projects and assets; the limited availability and accessibility of data, information, and tools for water management decision-making and performance evaluation; the fragmentation, heterogeneity, and compatibility of water governance and management structures, processes, and systems; and the need for more vital collaboration, coordination, and communication between various actors and sectors involved in smart water management.

2.2 Practical Advantages of Water 4.0

In 1994, VITUKI carried out a world survey on how computer technology is used in water and sewerage utilities in different countries [8]. The main areas identified were geographic information system (GIS) applications, process management of operations, water tariff billing and the implementation of expert systems. Even today a small Hungarian water utility is often lagging behind the French practice described two and a half decades ago. However, such development is often not an organic market development but a government decision. For example, in terms of spatial information technology, the process of electronic utility registration and reconciliation is well known in the industry. The authority builds its database by requesting data in electronic form, which significantly simplifies the process of utility planning and operation. Nevertheless, beyond this, given the correct data (e.g., from the smart meters below), web-based knowledge base systems can be built to facilitate the task of civil engineers in the construction of waterworks infrastructure [9].

The complexity of water and wastewater treatment plant management processes requires the use of many sensors and actuators. Digitalization can be used in the following processes:

1) Quantification and optimization of raw water intake
2) Energy efficiency in pump operation
3) Monitoring of reservoir water quality
4) Quantification of water to be treated and analysis of its quality
5) Optimization of the water treatment process (chemical, chlorine gas usealumina dosing).
6) It could be possible to determine the optimum pumping capacity at the point of delivery to the network, monitoring of the pressure of the water delivered to the network
7) Detection of pipe bursts and leaks
8) Real-time assessment of consumer demand
9) Pressure boosting in tower blocks
10) Protection of waterworks infrastructure

Despite having numerous data sources, there are many elements of the hydraulic network which cannot be measured. For those, we might implement virtual sensors to form a real-time dynamic hydraulic model for water loss reduction [10].

The operator can optimize the operation of pumps in pressurized sewerage systems, assess odor effects at sewage treatment plants, and optimize sewage treatment process flow. The utility could optimize the operation of wastewater treatment plants, streamline the sludge treatment process, increase the energy efficiency of biogas use, and measure the quality of treated wastewater before discharge to the intake [11]. Also, sludge storage can be optimized, and in the case of combined sewer systems, any bypassing processes can be scheduled. These processes can be combined with precision irrigation systems [12] to achieve a more efficient and optimally distributed load on the aquifer or water base.

The lowest-hanging fruit of digitalization is the use of smart meters, for which numerous technologies are available in the European market today. The smart water meter can be installed in the maintenance hole or basement, which can communicate with the water utility's systems from a place previously inaccessible by any other radio means, at low speed but with high reliability, using NB-IoT technology provided by the telecommunications operator [13]. Smart meters allow real-time analysis of water consumption patterns and on-demand adjustment of water treatment and supply technology to meet actual demand [14], making the water billing process more efficient than traditional solutions, as the bill can always be based on actual consumption data. The supplier can impose water consumption restrictions on the customer without on-site intervention. Data can be accessed directly from the

operational process in a billing system integrated with the ERP system. Postage costs and bill issuing logistics are minimized by issuing paperless but authentic bills sent in electronically signed PDFs.

The industry calls these solutions Water 4.0, Digital Water, Smart Water, and Internet of Water (IoW). The aforementioned solutions will undoubtedly reform municipal water management's entire infrastructure and operation and support more economically efficient operation of water utilities, customer satisfaction, environmental protection, process optimization, predictable maintenance and regulatory compliance. From a financial point of view, these solutions are characterized by lower operating costs (OPEX), better return on investment and higher revenues, while also increasing the company's economic value. Furthermore, in terms of operational security, higher levels of availability and more predictable human resource management can be achieved.

3 Security Considerations

To gain a deeper understanding of the cybersecurity challenges and opportunities in smart water management, this study did a literature review and secondary data analysis, focusing on academic articles, reports, and case studies related to cybersecurity incidents, trends, and measures in the water sector, both globally and in Hungary and Central Europe. The main findings and insights from this analysis are summarized below.

Digitalization has many positive benefits for individuals, organizations, and society, that we perceive day to day. However, we must not forget the dangers of being vulnerable to technology. In particular, the loss of IT applications and services, the automation of processes, and the difficulties arising from the complexity of systems must be taken into account.

3.1 Cybersecurity Threats

The new cybersecurity threats, vulnerabilities, and risks can significantly affect the safety, security, and reliability of critical water infrastructure and services and public health, the environment, and the economy. Some of the top cybersecurity challenges and threats in smart water management are shown below.

3.1.1 Cyber-Physical Attacks

These are attacks that target physical components and automation processes of water systems, such as pumps, valves, treatment plants, and distribution networks, through the manipulation or disruption of their digital control systems, such as supervisory

control and data acquisition (SCADA) systems, distributed control systems (DCS), programmable logic controllers (PLCs), and remote terminal units (RTUs) [15]. Cyber-physical attacks may cause physical damage, malfunction, or failure of water infrastructure, leading to water service interruptions, contamination, leakage, or flooding, with potentially severe consequences for public health, the environment, and the economy.

3.1.2 Data Breaches

Data breaches are unauthorized access, disclosure, or theft of sensitive or confidential data, such as personal or business-related information, financial transactions, or operational data, stored or transmitted in water management systems, such as customer billing systems, asset management systems, or monitoring and control systems [15]. Data breaches can result in privacy violations, legal consequences, identity theft, financial fraud, or industrial espionage, undermining the trust, reputation and compliance of water utilities and authorities.

3.1.3 Denial of Service (DoS) Attacks

These attacks aim to disrupt, overwhelm, or minimize the availability, performance, or functionality of water management systems, networks, or services, by flooding them with excessive data, requests, or traffic or exploiting their vulnerabilities and weaknesses or design flaws [15]. DoS attacks can cause temporary or permanent loss of access, control, or communication in water management and increase the workload, costs, and delays of system recovery and restoration.

As in the 2003 blackout mentioned in section 3.2, Cybersecurity incidents and trends in the water sector, water utility services can easily be lost in the event of the failure of some other critical infrastructure element. For example, without electricity, the pumps do not operate; therefore, drinking water will not be available. Nevertheless, even the loss of internet service could also be a problem for water utilities with a loss of communication, remote controls and a lack of data acquisition. The cause of these problems can be traced back to system unreliability and natural phenomena, but, most typically, human beings pose the greatest threat. This person could be an external attacker who wants to disrupt the industrial control systems through cybercrime or cyberwarfare. Industrial automation and control systems are also targets of military cyber operations, for example, with Advanced Persistent Threats (APT) [16].

3.1.4 Insider Threats

These are cybersecurity risks that originate from within the organization, such as employees, contractors, or partners, who have legitimate access and knowledge of water management systems, policies, or procedures, but who intentionally or unintentionally (e.g., error) misuse, abuse, or compromise them due to negligence, greed, revenge, or other motives [15]. Insider threats can be challenging to detect,

prevent, and respond to. They often involve exploiting trust, authority, or familiarity and can bypass or circumvent security controls and measures.

Internal staff also poses a threat by negligence or malicious acts. This includes using default passwords in ICS systems, which seems to be a trivial mistake, but still happens frequently because of the fictitious detachment of ICS systems from the Internet.

It is, therefore of paramount importance to support the digitization process with appropriate cybersecurity research and the resulting implementation of necessary information security control measures. IT-related areas need to employ staff with a higher level of cybersecurity expertise, which is generally not available in water and sewerage utilities.

3.1.5 Supply Chain Risks

These are cybersecurity risks that arise from the interdependencies and relationships between different actors and components in the supply chain of water management systems. These actors can be the vendors, suppliers, integrators, operators, and maintainers, who provide, produce, assemble, install, maintain, or update hardware, software, firmware, or services, which may contain vulnerabilities, defects, or backdoors, or which may be subject to tampering, sabotage, or counterfeit [15]. As process operation is highly dependent on the functionality of sensors, actuators, and the control system, the likelihood of downtime due to technical reasons is higher than in traditional systems [17]. The Internet of Things (IoT), the cloud computing infrastructure, and data processing are also major risk-increasing factors [18], which can only be effectively mitigated by a complex, interconnected control system [19]. Moreover, the failure of one system element can easily lead to a complete service outage. Supply chain risks can compromise the security, quality, and reliability of water management systems and create opportunities for attackers to exploit or manipulate these systems through various entry points, stages, or channels.

3.2 Cybersecurity Incidents and Trends in the Water Sector

Several high-profile cybersecurity incidents have been reported in the water sector in recent years, demonstrating the vulnerability and impact of water systems to cyber threats and attacks. An early example is the blackout that affected the northeastern United States and central Canada in August 2003 [20]. The world's second-largest blackout affected 55 million people and meant an almost total blackout for two days and abnormal service levels for two weeks. The exciting thing for us is that the incident started with a software failure at a service provider and the failure in an otherwise redundant system spread to other systems and took out the connected systems as well. For example, in 2016, a water treatment plant in the United States was compromised by a cyber-physical attack, which manipulated the levels of

disinfection chemicals used in the water treatment process, potentially endangering public health and safety [15]. According to the Israeli government, Palestinian hackers attempted to cause the mass chlorine poisoning at an Israeli water plant by switching the chlorine distribution system in April 2020, but the attack was hindered. The Palestinian side denied the accusation [21]. Other sources claim that the Safety Instrumented System (SIS), which is operated alongside the control system, averted the attack. Incidentally, SIS is also a control system design requirement.

Recent examples in the Russia-Ukraine armed conflict showed us multiple examples. During the conflict, many hacker groups have been attacking ICS/SCADA systems on both sides, as shown in Figures 2 and 3. Water utilities are attractive targets, as despite being critical infrastructure, the level of security protection is usually lower than electricity provision.

These incidents show the need for increased vigilance, preparedness, and resilience in the water sector against cybersecurity threats and risks.

The literature review and secondary data analysis revealed several emerging cybersecurity trends in the water sector, such as the growing number of reported cyber incidents, the increasing sophistication and complexity of cyber threats (like usage APTs), the expanding attack surface and vectors due to the proliferation of IoT devices, the evolving regulatory and compliance landscape, and the rising awareness and investment in cybersecurity measures and best practices. These trends indicate that cybersecurity is becoming a critical concern and priority for water utilities, authorities, and stakeholders, as well as for researchers, developers, and providers of smart water technologies and solutions.

Figure 2
Unverifiable information of a successful attack against a water supply SCADA system on 2022-03-07 reported by CyberThreat.Report [22]

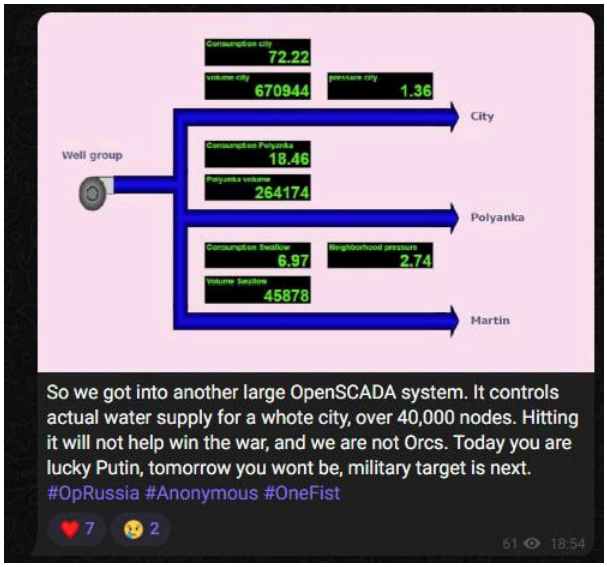


Figure 3
Unverifiable information of a successful attack against a water supply SCADA system on 2022-06-02 reported by CyberThreat.Report [22]

3.3 Cybersecurity Measures in the Water Sector

The research identified several cybersecurity measures and best practices that have been proposed or implemented in the water sector to address and mitigate information security risks and vulnerabilities. These measures include risk assessments, audits, and certifications, which help to identify, evaluate, and prioritize cybersecurity risks and controls; security policies, procedures, and guidelines, which define and communicate cybersecurity roles, responsibilities, and expectations; security training, education, and awareness training, which develop and maintain cybersecurity skills, knowledge, and culture; technical controls, such as firewalls, antivirus, encryption, and access controls, which protect and monitor information systems and networks; and incident response, recovery, and continuity plans, which prepare and guide organizations in handling and recovering from cybersecurity incidents [15]. In order to minimize the risk of cyber-physical attacks, we shall determine the critical objects, to strengthen their protection [23]. Most of the risks and countermeasures in computer networks and systems apply to industrial control systems, but they are biased. Access control, i.e., the restriction of access to authorized subjects, is basic in IT security, but it is more problematic in ICS. The reason of this is the lack of access control capabilities in many ICS devices. Against the insider threats, the solution could be a well-designed access control strategy [24]. The updates patching software flaws in ICS are rare and irregular in contrast to IT systems. Traditionally in ICS, functional safety is a priority over any other aspects. Patching of a running well-configured

industrial control system is apparently a risk from the aspect of functional safety, just like in IT systems. Although, in the IT systems there is a well-designed patch control process with central management (e.g., Windows Server Update Services, WSUS), rollback functions, test systems and patch piloting groups. Even software development and operation practices, like Data Execution Prevention (DEP), Address Space Layout Randomization (ASLR), Enhanced Mitigation Experience Toolkit (EMET), and Control Flow Enforcement (CFE) might be applied to minimize the risk of exploiting software vulnerabilities. [25] In the ICS systems, those methods and tools do not exist. Furthermore, in time-critical applications, neither a reboot is allowed. Network-based protection is by default based on air gaps, thus physical disconnection of networks. Historically there was no direct connection between the IT network and the operation technology (OT) network, which handles the ICT systems. Therefore, the network protection was unnecessary. The network-related problems arise when a previously air-gapped network is not just IT-network, but Internet-connected, because of industry 4.0 features [26]. Also, IT network security technologies might be only partly usable on the OT network. As functions are mostly time-critical, there is no delay allowed, so stateful packet inspection or application-level firewalls might not be used. Intrusion Prevention Systems (IPS) might disrupt regular communication when dropping packets, so they might be used with special caution.

4 Discussion: Implications and Recommendations

Based on the findings and insights from the literature review and secondary data analysis, as well as the context and conditions of Hungary and Central Europe, the following implications and recommendations can be derived for enhancing the cybersecurity and resilience of smart water management on the national level, or even in the region:

4.1 Strategies, Policies, Governance and Cooperation

Governments and authorities in Hungary and Central Europe should develop and implement comprehensive, coherent, and coordinated cybersecurity strategies and policies (or apply the current ones to the water sector), which address the specific risks, challenges, and opportunities of smart water management, and which are aligned with and supported by international, national, and sectoral cybersecurity frameworks, standards, and guidelines.

Some components of municipal water management are de facto critical infrastructures in Hungary, as specified in Act CLXVI of 2012 on the Identification, Designation and Protection of Critical Systems and Facilities [27] and in Government Decree 541/2013 (XII. 30.) on the Critical Water Management System Elements and Water Facilities

in Hungary [28]. Therefore, water supply is in the public interest, and wastewater collection and treatment, as a public health issue, shall also be continuously provided.

Freshwater is a strategic issue, and in many geographical regions (e.g., Israel, Tibet, Kashmir, Nile Valley), the use of limited freshwater resources is currently causing armed conflicts [29]. With water scarcity, the country is highly vulnerable to upstream countries if it does not have adequate storage capacity. This is a current problem in most of the Central European countries, mainly for social reasons, i.e., land use. Storage is important from both an economic (protection costs, utilization) and a strategic (indirect security risks) point of view [30].

Water utilities, authorities, and stakeholders in Hungary and Central Europe should establish and strengthen their cybersecurity governance and capacity by assigning clear organizational roles, responsibilities, and resources for cybersecurity management, integrating cybersecurity considerations into their organizational culture, processes, and systems, and by engaging in continuous learning, improvement, and innovation in cybersecurity practices and technologies [15].

Governments, utilities, industries, and research institutions in Hungary and Central Europe should promote and facilitate cybersecurity collaboration and information sharing in the water sector through various mechanisms and platforms, such as partnerships, networks, conferences, workshops, and publications, which enable the exchange of knowledge, experience, and intelligence about cybersecurity threats, vulnerabilities, incidents, and countermeasures, as well as the identification and dissemination of best practices and lessons learned in cybersecurity management and research [31].

Professional processes were initiated to increase legislative control concerning the electricity supply and distribution infrastructure, partly because of previous incidents and partly because of the higher level of IT exposure. This legislation started a decade ago in the United States, but just in recent years in Europe. However, as it is coupled with legislative intent, we can expect an increased resilience of the sector to cyberattacks [32].

As this is no longer the sole problem of countries, the European Union and NATO are dealing with this problem with different approaches. NATO addresses the strategic aspect and urges the preparation of member states for cyber operations as a distinct dimension of warfare [33]The EU deals more with the operative direction: in the Cybersecurity Act [34], one exact application is the information security-related certification of ICS components, as written in the Recommendations for the Implementation of the Industrial Automation & Control Systems Components Cybersecurity Certification Scheme (ICCS) by the European Commission, Joint Research Centre [35].

Since the digitalization described above is still in its infancy in Central Europe, it is advisable to use a Western European model for studies in this direction. A good example can be the German market, which also has many small public utilities.

German regulators are already addressing the cyber security of water utilities, with the KRITIS strategy and the resulting legislation setting out cyber security responsibilities for the entire German critical infrastructure [36].

In Hungary, a pioneering initiative is the SeConSys, a group of professionals and regulators working together on the cyber security of critical energy infrastructure. The Cybersecurity Handbook on Industrial Control Systems for Electricity [37], produced within this framework, is a missing piece of work that provides guidance for developing protection at the regulatory, design, and operational levels. In addition, case studies can be found in the Hungarian academic literature on the design of adequate cyber protection in the power sector [38].

4.2 Research, Development, and Innovation

Governments, public utilities, industries, and research institutions in Hungary and Central Europe should encourage and support cybersecurity research, development, and innovation in the water sector by funding (or applying to) multidisciplinary and collaborative research projects, programs, and centers, which focus on the design, evaluation, and deployment of secure, resilient, and trustworthy smart water technologies, systems, and services, as well as on the development and application of advanced cybersecurity methods, tools, and techniques, such as risk assessment, threat modeling, vulnerability analysis, intrusion detection, and incident response [15].

Although lagging behind, the maturity of water utilities' management and IT systems and the resulting cybersecurity risks will follow the same path as that of the electricity sector. The advance of digitalization in this area is inevitable, which also involves the emergence of new types of risks. Furthermore, the current unstable geopolitical situation makes critical infrastructure protection even more important. In order to prepare operators, legislators, and regulators for the challenges involved, it is essential to set up European professional and scientific initiatives to research the issue extensively.

University research has also started in this area. For example, the CYBERWATER NATO Advanced Research Workshop conducted research under the leadership of Harsha Ratnaweera, professor at the Norwegian University of Life Sciences [39].

Conclusions

This paper analyzed the increasing importance of smart water management, within the context of the fourth industrial revolution and its associated cybersecurity risks, particularly in Hungary and Central Europe. By examining the current state of smart water management and analyzing various cybersecurity threats, vulnerabilities, and risks in this critical sector, the study has sought to raise awareness about the need for enhanced security measures and practices in smart water management, as well as to provide a basis and course for future research.

The research was limited to the literature and the sources listed in the references. Even the application example, was based on one water utility, for which, the author had operational details. A broader research with more utilities (including large ones) could strengthen the findings.

Future research directions in cybersecurity and smart water management may include the following:

- Conducting empirical and comparative studies on the effectiveness and efficiency of different cybersecurity measures, practices, and technologies in the water sector.
- Exploring the ethical, legal, and social implications and challenges of cybersecurity and privacy in smart water management, such as, data ownership, consent and accountability.

By addressing these and other research questions and gaps, the academic and professional communities can contribute to the advancement of knowledge and practice, in cybersecurity and smart water management. This will help to ensure that the benefits and opportunities of digital transformation and innovation, in the water supply sector, are realized and sustained, without compromising the security, resilience and trust of critical water infrastructure/services, in Hungary, Central Europe and beyond.

Acknowledgement

The author would like to thank the two anonymous reviewers for their valuable feedback and suggestions, which helped to improve and extend the paper.

The research was supported by the Hungarian Academy of Sciences Bolyai János Research Scholarship (Grant No. BO/00372/22/9).

This research was supported by the ERDF project “CyberSecurity, CyberCrime and Critical Information Infrastructures Center of Excellence” (No. CZ.02.1.01 / 0.0 / 0.0 / 16_019 / 0000822).

References

[1] C. Roser, “AllAboutLean.com,” 2021. https://www.allaboutlean.com/industry-4-0/industry-4-0-2/ (accessed May 06, 2023).

[2] D. Wienecke-Janz, *Die Chronik der Deutschen (The Chronicle of the Germans)*. Gütersloh-Munich: Chronik Verlag, 2007.

[3] K. Schwab, *The Fourth Industrial Revolution*. Crown Business, 2017.

[4] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, “Industry 4.0,” *Business & Information Systems Engineering*, Vol. 6, No. 4, pp. 239–242, 2014, doi: 10.1007/s12599-014-0334-4.

[5] A. D. Gupta, P. Pandey, A. Feijóo, Z. M. Yaseen, and N. D. Bokde, "Smart Water Technology for Efficient Water Resource Management: A Review," *Energies (Basel)*, Vol. 13, No. 23, p. 6268, Nov. 2020, doi: 10.3390/en13236268.

[6] M. Makai, "Wastewater sector issues after the KEHOP, before the end of the derogation. Strengthening higher education in water management in the framework of smart specialisation," Decentralized Wastewater Treatment Conference at National University of Public Service, Faculty of Water Sciences, Baja, Mar. 21, 2019.

[7] S. R. Krishnan *et al.*, "Smart Water Resource Management Using Artificial Intelligence—A Review," *Sustainability*, Vol. 14, No. 20, p. 13384, Oct. 2022, doi: 10.3390/su142013384.

[8] J. Deri, *Computerised technologies for water and sewerage plants (in Hungarian: Számítógépesített technológiák víz- és csatornaműveknél)*. Budapest: Vituki Innosystem, 1994.

[9] R. A. Stewart, R. Willis, D. Giurco, K. Panuwatwanich, and G. Capati, "Web-based knowledge management system: linking smart metering to the future of urban water planning," *Australian Planner*, Vol. 47, No. 2, pp. 66–74, Jun. 2010, doi: 10.1080/07293681003767769.

[10] M. S. Osman, A. M. Abu-Mahfouz, and P. R. Page, "A Survey on Data Imputation Techniques: Water Distribution System as a Use Case," *IEEE Access*, Vol. 6, pp. 63279–63291, 2018, doi: 10.1109/ACCESS.2018.2877269.

[11] M. Patziger, *Efficient operation of medium and small wastewater treatment plants (in Hungarian: Közepes és kis szennyvíztisztító telepek hatékony üzemeltetése)*. Budapest: Hungarian Water Utility Association, 2018.

[12] S. Takács, T. Bíró, L. Helyes, and Z. Pék, "Variable rate precision irrigation technology for deficit irrigation of processing tomato," Vol. 68, pp. 234–244, 2019, doi: 10.1002/ird.2299.

[13] S. Alvisi *et al.*, "Wireless Middleware Solutions for Smart Water Metering," *Sensors*, Vol. 19, No. 8, p. 1853, Apr. 2019, doi: 10.3390/s19081853.

[14] S. D. B. Moraes, C. Langhi, and M. Crivelaro, "How an existing telecommunications network can support the deployment of smart meters in a water utility?," *Independent Journal of Management & Production*, Vol. 6, No. 4, 2015, doi: 10.14807/ijmp.v6i4.351.

[15] Y. Cherdantseva *et al.*, "A review of cyber security risk assessment methods for SCADA systems," *Comput Secur*, Vol. 56, pp. 1–27, Feb. 2016, doi: 10.1016/J.COSE.2015.09.009.

[16] W. Steingartner and D. Galinec, "Cyber Threats and Cyber Deception in Hybrid Warfare," *Acta Polytechnica Hungarica*, Vol. 18, No. 3, pp. 25–45, 2021, doi: 10.12700/APH.18.3.2021.3.2.

[17] M. Moy de Vitry, M. Y. Schneider, O. Wani, L. Manny, J. P. Leitão, and S. Eggimann, "Smart urban water systems: what could possibly go wrong?," *Environmental Research Letters*, Vol. 14, No. 8, p. 081001, Aug. 2019, doi: 10.1088/1748-9326/ab3761.

[18] R. O. Andrade, S. G. Yoo, L. Tello-Oquendo, and I. Ortiz-Garcés, "A Comprehensive Study of the IoT Cybersecurity in Smart Cities," *IEEE Access*, Vol. 8, pp. 228922–228941, 2020, doi: 10.1109/ACCESS.2020.3046442.

[19] A. Toth, "Cloud of Things Security Challenges and Solutions; Cloud of Things Security Challenges and Solutions," in *Communication and Information Technologies (KIT)*, Vysoke Tatry, Slovakia, 2021, pp. 1–6. doi: 10.1109/KIT52904.2021.9583760.

[20] M. Tabibzadeh and S. Mirzaei, "A system-oriented framework for risk and resiliency analysis of power blackouts," in *Proceedings of the 2016 Industrial and Systems Engineering Research Conference, ISERC 2016*, 2020, pp. 582–587.

[21] "Iranian Cyberattack Aimed to Raise Chlorine Level in Israeli Water, Report Says," *Haaretz*, Jun. 01, 2020. Accessed: Jun. 04, 2022. [Online]. Available: https://www.haaretz.com/israel-news/iranian-cyberattack-aimed-to-raise-chlorine-level-in-israeli-water-report-says-1.8886235

[22] CyEx Kft, "Cyberthreat.Report closed Facebook group," Mar. 07, 2022. https://www.facebook.com/groups/469363908185824 (accessed May 08, 2023).

[23] A. Massel and D. Gaskova, "Identification of Critical Objects in Reliance on Cyber Threats in the Energy Sector," *Acta Polytechnica Hungarica*, Vol. 17, No. 8, pp. 61–73, 2020, doi: 10.12700/APH.17.8.2020.8.5.

[24] B. Leander, A. Čaušević, H. Hansson, and T. Lindström, "Toward an Ideal Access Control Strategy for Industry 4.0 Manufacturing Systems," *IEEE Access*, Vol. 9, pp. 114037–114050, 2021, doi: 10.1109/ACCESS.2021.3104649.

[25] L. Erdődi and A. Jøsang, "Exploitation vs. Prevention: The ongoing saga of software vulnerabilities," *Acta Polytechnica Hungarica*, Vol. 17, No. 7, pp. 199–218, 2020, doi: 10.12700/APH.17.7.2020.7.11.

[26] E. D. Knapp and J. T. Langill, *Industrial Network Security: Securing Critical Infrastructure Networks for Smart Grid, SCADA, and Other Industrial Control Systems*. Waltham, MA: Syngress, 2015.

[27] Hungarian Parliament, *Act on Identification, Designation and Protection of Critical Systems and Facilities*. 2012. Accessed: May 08, 2023. [Online]. Available: https://njt.hu/eli/TV/2012/166

[28] Hungarian Government, *Government Decree 541/2013 (XII. 30.) on the Critical Water Management System Elements and Water Facilities*. Hungary, 2013.

[29] J. Padányi, "Water conflicts," *Military Science*, Vol. 25, No. e, pp. 272–284, 2015.

[30] K. Ligetvári, "Water security problems in hungary in comparison with EU and world tendencies," *Military Science*, Vol. 23, No. 1, pp. 4–13, 2013.

[31] Z. Bederna and Z. Rajnai, "Analysis of the cybersecurity ecosystem in the European Union," *International Cybersecurity Law Review*, Vol. 3, No. 1, pp. 35–49, Jun. 2022, doi: 10.1365/s43439-022-00048-9.

[32] E. D. Knapp and R. Samani, *Applied Cyber Security and the Smart Grid: Implementing Security Controls into the Modern Power Infrastructure*. Waltham, MA: Syngress, 2013.

[33] P. Bányász, C. Krasznay, and A. Tóth, "NATO's cybersecurity politics (in Hungarian: A NATO kibervédelmi szakpolitikája)," in *Today's NATO: Status and roles of the alliance (in Hungarian: A mai NATO: A szövetség helyzete és feladatai)*, Z. Szenes, Ed., Budapest: HM Zrínyi, 2021, pp. 130–149.

[34] European Parliament and the Council, *Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act)*. European Union, 2019. Accessed: May 08, 2023. [Online]. Available: http://data.europa.eu/eli/reg/2019/881/oj

[35] P. ; Theron *et al.*, "Recommendations for the Implementation of the Industrial Automation & Control Systems Components Cybersecurity Certification Scheme (ICCS)," Ispra, 2020. Accessed: Oct. 01, 2022. [Online]. Available: https://ec.europa.eu/jrc

[36] J. Fettig and M. Oldenburg, "Overview: Preparedness in the Water Supply and the Sanitation and Sewerage Sectors in Germany and Europe," in *Physical and Cyber Safety in Critical Water Infrastructure*, H. Ratnaweera and O. A. Pivovarov, Eds., Amsterdam: IOS Press, 2019.

[37] I. Angyal *et al.*, *Cybersecurity handbook on industrial control systems for electricity service provision*, 2nd ed. Budapest: National Cyber Defense

Institute of the National Security Service, 2021. [Online]. Available: https://m2.mtmt.hu/api/publication/32462794

[38] C. Krasznay and M. Danyek, "Protecting the National Electricity System in the Cyberspace – A Case Study," in *Information- and cybersecurity*, B. Török, Ed., Budapest: Ludovika Egyetemi Kiadó, 2020, pp. 149–163.

[39] NATO, "NATO Newsroom," 2018. https://www.nato.int/cps/en/natohq/news_157806.htm (accessed May 06, 2023).

Analysis of the Soil Selective Potassium Content, using Multifrequency EC Sensors

János Horváth[1], László Kátai[2], István Szabó[2]

[1]Doctoral School of Mechanical Engineering, Hungarian University of Agriculture and Life Sciences, Páter Károly utca 1, 2100 Gödöllő, Hungary, Horvath.Janos.7@phd.uni-mate.hu

[2]Institute of Technology, Hungarian University of Agriculture and Life Sciences, Páter Károly utca 1, H-2100 Gödöllő, Hungary; Katai.Laszlo@uni-mate.hu; Szabo.Istvan.prof@uni-mate.hu

Abstract: The relative salinity of soil is an important issue in today's sustainable, site-specific Precision Agricultural practice. The need to replace salt has an enormous impact on production costs, regarding current input prices. The standard method for measuring the salinity of a soil, based on a laboratory test method, is expensive and cumbersome and only the total salinity content can be determined. Growers need to get data faster and cheaper. With the usage of on-the-go soil sensors, this process can be significantly improved. This article presents the first step of developing a quick, cost-saving and easy-to-use measurement method for determining the selective potassium content on the spot, in real time. In our work, we developed a new measurement and calculation method using the analytical application of multifrequency conductometry, to determine selective salinity. It is proved that the electrical conductivity (EC) develops according to different functions, by applying solutions containing K+ ions with a defined concentration, by saturating the soil with a continuously, but reasonably varying measuring frequency. Based on this recognition, the selective salinity of soil can only be determined by interpolation from a frequency-series of EC measurements, in soils containing only one electrolyte.

Keywords: digital agriculture; electrical conductivity; soil salinity; variable frequency; interpolation

1 Introduction

1.1 Digital Farming Concept

Nowadays, there is a constantly growing demand to know our soils as thoroughly as possible and the precision technology toolkits already present in our machines

provides effective support. Numerous smart solutions help farmers in all areas of Precision Agriculture, onboard and offboard the tractor.

Digital agriculture is still a young concept. Synonymous terms are agriculture 4.0, smart farming or precision farming, whereby the terms are to be considered evolutionary. The latter term has its origins back in the 90s. As Figure 1 shows, Griepentrog differentiates between the terms as follows [1]:

- **Precision Farming** focuses on site-specific processing in which the growth conditions are optimized by means of sensors and application technology
- **Smart Farming** adds a level of decision support through fusion and analysis of information
- **Digital Farming** (or Agriculture 4.0 or Farming 4.0) includes the Internet of Things as well as Cloud Computing and Big Data

Figure 1
The development of Precision to Digital Farming [1]

In the universe of digital agriculture, the basic goal of site-specific management of agricultural inputs is to increase the profitability of crop production by improving the quality of output and protecting our environment. The process of getting to know our soil as is the Farm Management Information System (FMIS) designed considering the benefits of data processing and data use provided by digitization, functions as a unit of 5 processes [2]:

- **Data acquisition**: measurement and recording of data on the field, part-fields
- **Data collection and storage**: collecting data on a specific field, a specific location within a field
- **Data processing**: providing useful information for decision-making
- **Decision-making**: defines the cultivation technology in detail, broken down for each operation, as well as the details of tillage, nutrient replenishment, planting and other operations, together with characteristic parameters and setup

- **Executing the operation**: turning the decision-making results into practice

As a first step of the recognition process, it is necessary to obtain data on the variability of the parameters describing the soil properties within the field for geo-positioning. With the correct and sensible use of sensor technology, farmers can better understand their crops and soils while conserving resources and at the same time saving resources and reducing their impact on the environment [3]. If we do not have access to the necessary and relevant data, then the decision-making and the performance of the operation are affected by the quantity and quality of the collected data, which determines the intervention and the cost-effectiveness of the intervention on the field.

1.2 Testing Soil Properties with Sensors

Today, sensor technology is one of the fastest growing areas of technology. A sensor is a device that can detect a change in the physical or chemical environment, which then converts it into electrical signals, both electric current and voltage. The precision agriculture mainly relies on real-time monitoring of soil conditions using information technology and (Global Position System) GPS technology, and then analyzing and managing the spatial-temporal variability of soil and field crops. These information help make decision on precision application of crop inputs including water and fertilizer. It can improve the efficiency and reduce the losses of water and fertilizer [4].

The ever-increasing prices of fertilizer and growing ecological concern over chemical run-off into sources of drinking water, have brought the issues of precision agriculture and site-specific management to the forefront of present-day technological development within agriculture and ecology. Due to increases in the cost of fertilizer production inputs—predominantly nitrate (N), phosphate (P), and potassium (K) those in agriculture are looking for ways to optimize plant yield while minimizing the application and consumption of fertilizer. Since these macro-nutrients vary even on a small scale throughout a cultivated field, numerous researchers have attempted to develop an on-the-go sensing apparatus that can map the presence of these chemicals in situ so that this map, once overlaid with parameters such as pH, electrical conductivity (EC), crop yield, and mechanical properties of the soil, can give a precise spatially varying prescription for fertilizer application [5].

The global Agricultural Sensors market was valued at USD 1,505.4 million in 2020 and expected to reach USD 3200.8 million by the year 2028, at a Compound Annual Growth Rate (CAGR) of 11.04%. Smart Sensors allow farmers to maximize yields using minimal reserves such as fertilizer, water, and seeds. By utilizing sensors and mapping fields, farmers can commence realizing their crops at a micro-scale, conserve resources, and lessen influences on the ecosystem. Smart agriculture has

roots moving back to the 1980s when the Global Positioning System (GPS) capability became accessible for civilian use. Once farmers were able to map their crop fields accurately, they could monitor and use fertilizer and weed treatments only to parts that required it. During the 1990s, early precision agriculture users implemented crop yield monitoring to create fertilizer and pH correction suggestions. As more variables could be calculated and entered into a crop model, more accurate recommendations for fertilizer application, watering, and even peak yield harvesting could be made. The device capture analyzes and transmits information like temperature, humidity, pressure, water content, etc. using radio signals. It gathers the data and sends it to the base station. The base station then analyses the data and carries it for further processing. Usually, sensor networks have a base station known as sink and several other sensors too, which sense and transmit the signals along with sending information to other nodes. Weather stations are self-contained units placed at various sites throughout growing fields. These stations have a mixture of sensors suitable for the local crops and climate. Data such as air temperature, soil temperature at various depths, dew point temperature, wind direction, relative humidity, rainfall, leaf wetness, chlorophyll, wind speed, solar radiation, and atmospheric pressure are measured and recorded at predetermined intervals [6].

For sensor measurement, the following sensor groups can be found in crop production: crop sensors, environment sensors, function monitoring sensors and soil sensors. From the growth of forecast solutions using sensors, it is very easy to recognize that the future of site-specific crop production is moving towards a sensor-based approach because obtaining important soil characteristics quickly and cheaply is still one of the biggest challenges in precision agriculture today. A number of researchers and manufacturers are trying to develop on-the-go soil sensors, to directly measure the mechanical, physical, and chemical properties of the soil. The disadvantage of the practical use of the increasingly widespread ground sensors is that they are less accurate than individual sampling and laboratory tests, but the advantage is that they are suitable for rapid measurement and are therefore cheaper in practice [7]. In the future, mechanized soil testing and nutrient mapping solutions will become widely available using faster and more cost-effective measurement tools. To implement sustainable agricultural and environmental management, a better understanding of the soil at increasingly finer scales is needed. Conventional soil sampling and laboratory analyses cannot provide this information because they are slow and expensive [8].

1.3. Proximal Topsoil Sensors

Proximity detection is defined as using sensors applied in the field to obtain signals from the ground with the sensor unit in contact with or near the ground (within 2 m) [9]. The following measurement principles and sensors are distinguished in practice [10]:

- Electrical Conductivity (EC) Sensors
- Ground Penetrating Radar and Reflectometers
- Visible (VIS) – Near-Infrared (NIR) – Mid-Infrared (MIR) Diffuse Reflectance Sensors
- Magnetic Susceptibility Sensors
- Gamma-Ray sensors
- X-ray Sensors
- Other Proximal Soil Sensors like photoacoustic spectroscopy, laser-induced breakdown spectroscopy laser-induced fluorescence spectroscopy and inelastic neutron scattering. Mechanical sensors that measure soil penetration resistance have also been used extensively and integrated with other sensors in soil science and precision agriculture, with applications ranging from soil compaction assessment to 3D modelling of soil layers.

Of course, the different sensors can be combined as desired, which is defined as using a multisensory measurement.

1.4 Conductometry - Measurement of Electrical Conductivity

Soil properties often vary significantly within a field, and one of the challenges in precision agriculture is collecting enough soil data to accurately delineate this variability. Soil electrical conductivity (EC) has become a widely used tool for mapping soil variability within fields. Soil EC measurements are typically correlated with soil texture, moisture, and salinity. Soil texture is an important factor for crop yields, because it relates to the water-holding capacity, cation-exchange capacity, rooting depths, drainage and other properties, that can impact crop production [11].

The good conductance of parts is characterized by electrical conductivity. Conductometry is an analytical method based on the measurement of the electrical conductivity of solutions. In analytical chemistry, the electrical conductivity of electrolyte solutions is measured by conductometric methods, and analytical information is derived from this and its changes due to chemical reactions. The conductivity of materials (symbol G, the unit is siemens, S) is the reciprocal of their electrical (ohmic) resistance (symbol R, the unit is ohm, Ω). Electrical conduction requires the presence of charge carriers (e.g., electrons or anions and cations) that are able to move under the action of the electric field. The ion migration in the solution due to the electric field is the reason for the electrical conduction of the solutions. Based on this, electrical conductors and insulators are to be distinguished. Pure (distilled) water, since it contains only very small charge carriers in the concentration $[H^+] = [OH^-] \approx 10^{-7}$ mol l^{-1} corresponding to autoproteolysis, conducts electricity only to a very small extent, therefore it can be

considered as an insulator. However, the concentration of cations and anions in aqueous solutions of electrolytes can be significant, making them mostly conductive depending on the degree of electrolytic dissociation. Each ion in the solution makes some contribution to the conduction value. These are inseparable, so conductometry is not suitable for the selective measurement of individual ions, i.e., it is not an ion-specific method [12]. Therefore, its analytical application is limited to the study of systems that

- Contain only a single electrolyte (so the contribution of the "background" is negligible)
- Chemical reactions take place in them, during which the mobility of the ions that make up the system changes significantly (compared to non-zero "background") [12]

1.5 Principles of Soil Electrical Conductivity

What about special solutions to sample the soil? Do conductivity measurements only provide information on all salinity in the soil and are they not suitable for ion-selective measurements? The measured conductivity is therefore only approximate information [13]. Figure 2 shows the possible directions of displacement of the charged particles during the 3-phase, unsaturated soil conductivity measurement. The air acts as an insulating medium [13] [14]. Three pathways of current flow contribute to the EC of a soil:

(i) A liquid phase pathway via dissolved solids contained in the soil water occupying the large pores

(ii) A solid–liquid phase pathway primarily via exchangeable cations associated with clay minerals

(iii) A solid pathway via soil particles that are in direct and continuous contact with one another [15].

These three pathways of current flow are illustrated in Figure 2, the air is shown in white, the liquid in gray, and the soil particles are scored [14].

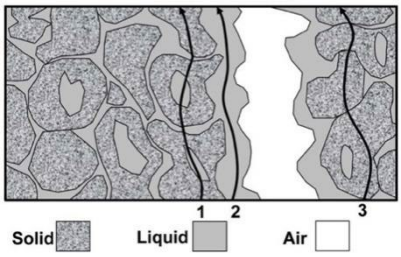


Figure 2
Direction of electrical conduction in 3-phase soil [13] [15]

Nevertheless, the explosive spread of EC measurements since the 90s can be attributed to 2 reasons. On the one hand, with the development of technology, the size of conductivity measuring devices has been significantly reduced and are also available as mobile devices, and on the other hand, satellite positioning systems have become a part of our lives and conductivity is an easily measurable feature.

1.6 Chemical Composition of Soil Moisture (Soil Solution)

Various inorganic salts, organic matter and gases are dissolved in the soil moisture. Dissolved mineral salts are dissociated into positively and negatively charged ions and the ions are surrounded by a hydrate shell. The following ions are mainly present in the soil solution:

- Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, NH^{+4} and in some soils Al^{3+}, Fe^{3+} or Fe^{2+} (cations)
- HCO^{3-}, CO_3^{2-}, Cl^{-}, SO_4^{2-}, NO^{3-}, H_2PO^{4-}, HPO_4^{2-} (anions)

Most of the dissolved organic compounds are organic acids and low molecular weight humic substances, with CO_2 and O_2 being the most significant dissolved gases. Soluble substances in the soil are mostly products of weathering and soil formation, but salts also enter the soil solution with groundwater close to the surface. In addition, fertilizers and compounds applied to irrigation water in agricultural areas modify the salinity of the soil [16].

1.7 Hydrated Ions

The interaction between solvated ions and water has been a subject of great interest due to its importance in various chemical, biological, and environmental processes [17]. Various experimental and theoretical studies have explored aqueous solutions of ions, both in the bulk [18] [19] and at interfaces [20-22].

The size of the hydrate shell in aqueous solution is illustrated in Figure 3. The smaller the diameter of the dehydrated cation and the larger its charge is, the thicker the hydrate sphere in aqueous solution is. This also affects the mobility of hydrated ions. Potassium is the most mobile element in the soil and as such readily available to plants however care must be taken to avoid it leaching out of the vadose zone of the soil because of its high mobility [23]. Explanation for conclusion:

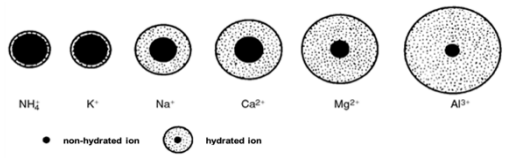


Figure 3
The size of the aqueous hydrate cover for each element in the soil [16]

In fact, the common cations in soil have a wide range of sizes, as shown in Fig. 4. Rengasamy and Sumner [24] showed that the dispersive effect of Na is greater than that of K, and that the flocculating effect of Ca is greater than that of Mg. Conventionally, these cations have been called base or base-forming cations [25]. Hydrated radii are critical to the composition of soil.

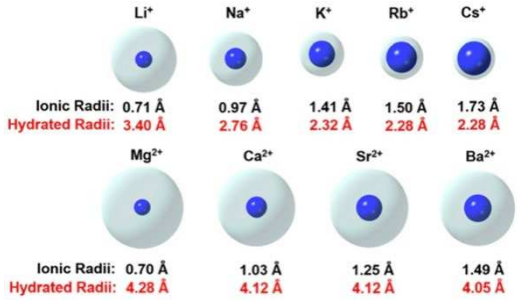


Figure 4

The ionic and hydrated radii (effective sizes) in angstroms, of some of the common ions [26]

1.7 Electrical Conductivity Sensors

Because soil salinization is one of the critical global problems threatening land productivity [27]. Saline soils have been reported from more than 100 countries around the world and covers more than 1125 million hectares of land [28]. It is increasing at about 1-2% every year and is estimated that 50% of the available arable lands can be affected by soil salinity by 2050 and would pose a serious threat to the sustainable development of global agriculture [29].

It is obvious that the measurement of electrical conductivity (EC) is one of the in-situ tools for soil testing required for precision farming and accordingly there is a large-scale literature on the development of EC detectors. These detectors typically and traditionally have 4 electrodes [30-32]. The traditional current-voltage four-electrode method to better fit for in-situ measuring and aimed at developing an in-situ soil EC detector with low price, easy operation, high measurement precision, integral control procedures and data processing procedures [30]. Apparent soil electrical conductivity is one of the simplest, least expensive soil measurements to obtain useful information about soil characteristics, which have a vital role in precision agriculture [31].

The aim of this paper is to present the basic research of measuring selective potassium content of the soil by multifrequency electrical conductivity measurements based on principles of conductometry (EC measurement), which is being developed within the framework of Hungarian University of Agriculture and Life Sciences, Institute of Technology. The methodology development focuses on

how to replace slower, more complex, and thus relatively expensive laboratory determination of salinity with as many on-the-go measurements as possible. It is noted that to the best of our current knowledge, conductivity tests with soil sensors alone are not sufficient to infer the salinity properties of a given field. The measurement is affected by the unequal distribution of different nutrients, different pH conditions, different particle compositions, differences in organic matter content or even temperature [13].

2 Methods and Materials

2.1 Specification of Solution

The measurements were performed in the Lab of Institute of Technology, in Hungarian University of Agriculture and Life Sciences. The laboratory temperature was set at a constant 22°C. The solutions and the soil were used for the measurement at 22°C.

During the measurement we wanted to study the effect of the cations of the dissolved salts in the soil moisture under natural conditions. The selected cation was K^+, because the reasonable replacement of the use of K^+ microelements in cultivated plants is a burning and expensive issue. We had to introduce these ions into the soil solution in the form of water-soluble salts. The salts and chlorides are very soluble in water, so we chose the chlorine salts of the K^+ cation for the experiment. We prepared 1 M concentration (1 M = 1 mol dm^{-3}) stock solutions from Potassium chloride, ACS Reagent, 99.0-100.5% CAS: 7447-40-7. These solutions were further diluted with the dilution series: 1 M; 0.66 M; 0.5 M; 0.33 M and distilled water i.e., 0 M was the reference measurement. Between the measurements we always washed the system with distilled water, so that any ions left on the electrode during the previous measurement, do not cause a measurement error.

2.2 Specification of Soil

As a model soil for our investigations, the same pH_{H2O} <7 soil has been chosen which had no detectable $CaCO_3$ content. The test soil was collected in Szárítópuszta in Gödöllő town, using 0-20 cm layers of arable land (47° 35' 47.65" N, 19° 21' 18.54" E). Soil type based on IUSS Working Group WRB, 2015 was Eutrict Arenosol (Aeolic, Aric, Ochric, Raptic) from elevation: 232 m with texture of sand.

2.3 Hardware Specification

Elements of the measuring circuit as in Figure 5:

1. SOURCETRONIC ST2829C Precision LCR meter with USB stick
2. SOURCETRONIC ST26011B Test Fixture: the resolution of the instrument is 0.00001 nS, the basic accuracy is 0.05 %. During the measurement, the voltage of the electrodes is 10 V DC. The output impedance of the instrument was 100 Ω
3. KSP-F01A Dosing Pump
4. Measuring cup (500 ml, 50 ml increments) to store the input solution
5. Measuring cylinder ∅ 61 mm x 137 mm
6. Measuring electrodes 2 pcs, ∅ 2 mm, length: 80 mm full length insulated, uninsulated part length 5 mm, probe distance 12 mm, material is stainless steel
7. Piping, inner ∅ 3 mm
8. Tray with raiser

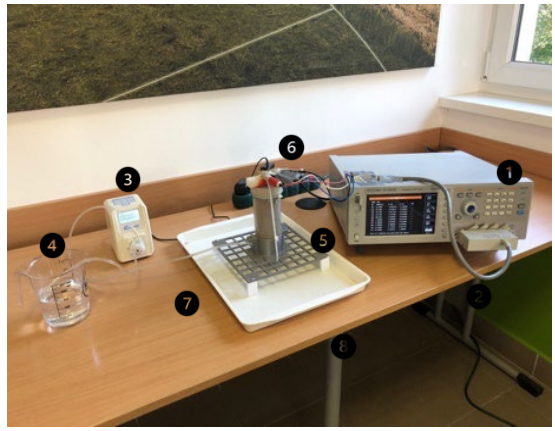


Figure 5
The measuring circuits

The LCR meter measured and recorded the electrical conductivity between the two stainless steel electrodes placed in the ground paste in an endless series at a predetermined program frequency and restarted the measurement in an endless cycle upon completion of the measurement sequence. The frequency values used for the measurements were recorded in the range of 20 Hz and 10^6 Hz.

Each new measurement was started by calibrating the peristaltic pump. The liquid delivery of the pump was set to 0.5 g s^{-1}. During the measurements, the liquid

solution was pushed into the cylinder (Figure 6) by a pump at the bottom of the measuring cylinder, where the solution was introduced into the cylinder through a distribution plug made of 3D printing so that the soil in the cylinder could saturate evenly. The solution left the measuring cylinder on the perforation formed at the top of the cylinder and collected on the tray under the measuring cylinder, from which the used output liquid was continuously emptied.

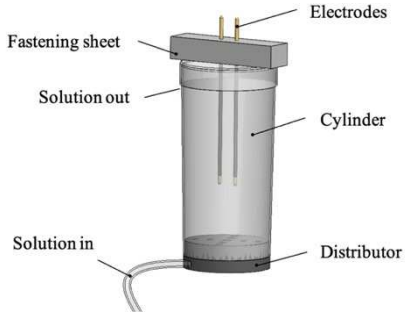


Figure 6
The measuring cylinder and connecting parts

Before starting each new measurement, we washed the soil paste with distilled water in the measuring cylinder so that the ions remaining after the previous measurement did not affect the measurement results. The experiments were repeated three times with each mixed input solution by adding 300 ml of the input solution and each filling was followed by neutralization with 300 ml of distilled water. If the soil paste was not reduced to less than 0.1 mS during the washing, an additional washing with 100 ml of distilled water was performed. Each measurement was repeated three times.

3 Results and Discussions

3.1 Reference Functions to Determine the K Content

From the reference measurements, we determined the arithmetic mean of the measured EC values for each concentration and frequency by examining the adjacent values and averaging them from the values where the change did not exceed 5% ($\Delta < 5\%$). Therefore, the frequency-dependent averages for a complete measurement were determined. After performing the calculations, it was possible to record the measurement results for the calculated data sets per ion, concentration and measurement frequency.

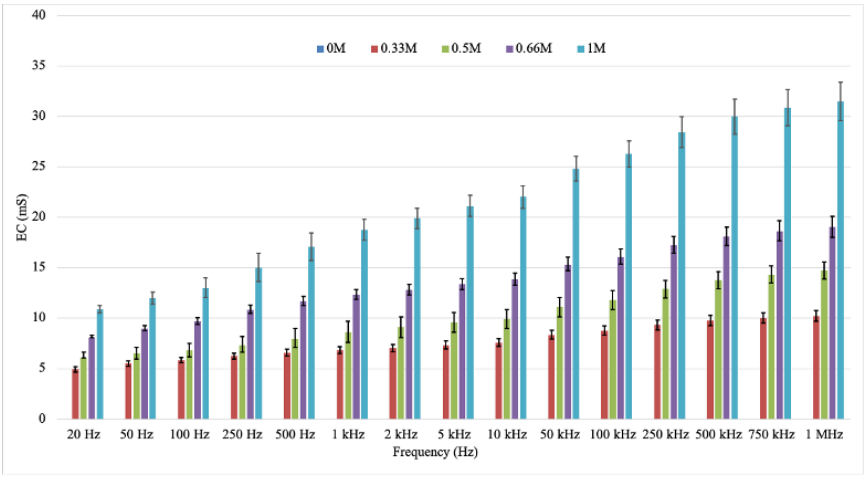


Figure 7

EC values for different concentrations of KCl solutions with standard deviation

When examining the quality of the derived data through the standard deviation of the mean estimate (Figure 7). It is much more informative to plot the change in EC of the solutions of the investigated concentration as a function of the logarithmic frequency. Figure 8 not only shows that the EC grows with increasing concentration, but it is also clear that if the measurement frequency is increased, the measured EC value also increases.

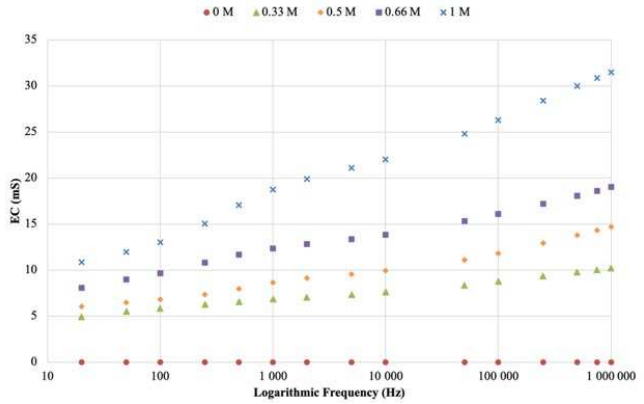


Figure 8

EC as a function of the logarithmic measurement frequency for different concentrations of KCl solutions

Plotting the measurements in a 3D space, the following diagram shown in Figure 9, can be drawn:

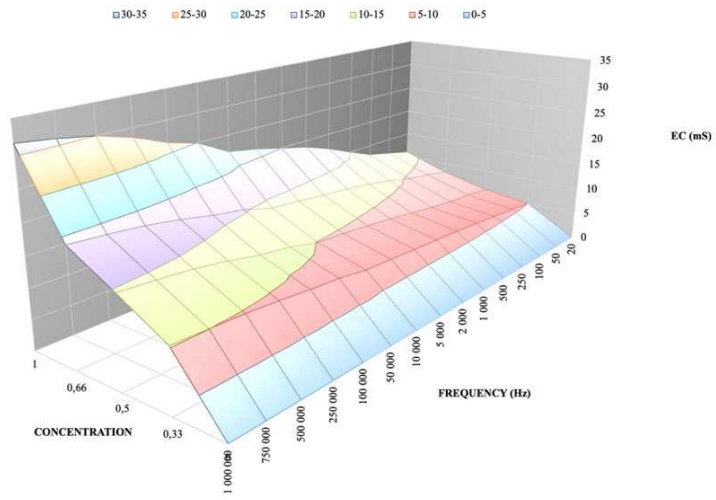


Figure 9
EC as a function of K^+ concentration and frequency

Finally, when plotting the frequency on a linear scale and plotting the EC measurement results of the solutions of different concentrations and fitting a function to the obtained points, considering that $R^2 > 0.9$. The reference functions of the different K^+ cation concentrations are illustrated in Figure 10.

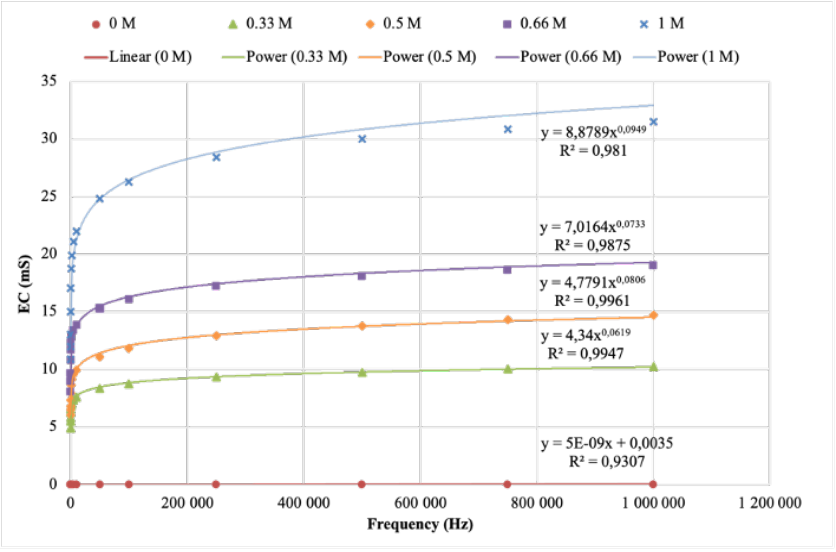


Figure 10
EC saturation curve for solutions with different K+ concentrations on a natural frequency scale

3.2 How to Determine the Concentration of an Unknown Solution Based on our Experiments

Relative EC is measured, but the selective salinity of the soil must be inferred. The solution requires interpolating the obtained measurement results. The solution is to determine the concentration by interpolating the reference values. Premises:

- The analytical application of conductometry is limited to the study of systems that contain only a single electrolyte, so the contribution of the "background" is negligible
- A complete measurement is a measurement performed over the entire frequency range
- It is necessary to use the same frequencies

3.3 The Application of the Following Notations for Deduction

- $x_1 < x_2 < ... < x_n$ the preliminary measurement frequencies
- $z_1 < z_2 < ... < z_n$ EC values for the measurement frequencies in a series of measurements

According to Figure 11 at the given frequency x_1 of the previously measured reference EC values, let $u_1$and v_1 be the reference values which enclose the measured value of z_1 i.e.:

$$u_1 \leq z_1 \leq v_1 \tag{1}$$

and so on, in x_n it is u_n and v_n that enclose the value of z_n, that is:

$$u_n \leq z_n \leq v_n \tag{2}$$

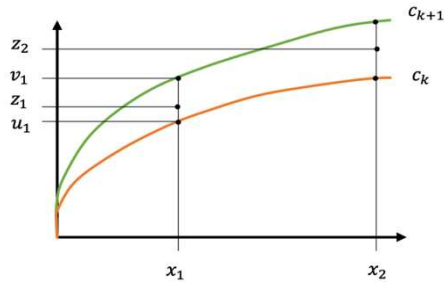


Figure 11
Determination of concentration by interpolation of reference values

Suppose that all points (x_i , z_i) fall between the same two adjacent EC curves.

Use:

$$i = 1, 2, \dots, n \text{ fixed} \tag{3}$$

and these two curves correspond to the concentrations c_k and c_{k+1}

We are looking for a $x_i \in [0,1]$ value, to which:

$$z_i = x_i u_i + (1 - x_i) v_i \tag{4}$$

from

$$x_i = \frac{v_i - z_n}{v_i - n_i}, \text{ where} \tag{5}$$

$$i = 1, 2, \dots, n \tag{6}$$

x_i is the average of interpolation constants

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{7}$$

This interpolates between c_k and c_{k+1} reference concentrations:

$$\bar{c} = \bar{x} c_k + (1 - \bar{x}) c_{k+1} \tag{8}$$

Ergo, the estimated concentration of the unknown solution measured will be the average of the concentrations that interpolated between the EC values given by the reference measurements at the measurement frequency occurring in one measurement.

Conclusions

With the current development of soil sensors, it becomes possible to measure soil properties in real-time, so growers can now get immediate information about the state of the most important resource, the soil, which can save resources and reduce the impact on the environment. The use of soil sensors enables rapid, immediate and cost saving soil testing [33] [34] and nutrient mapping solutions.

In our current research, we were looking for the answers to whether one of the simplest and cheapest measurement methods, i.e., the new approach to measuring the electrical conductivity as a parameter of the soil, offers the possibility to measure selective salinity in the laboratory. Our experiments demonstrated that there is a significant correlation between the salt concentration of the soil, the measurement frequency and the measured EC. As the frequency has been increased, the measured EC of the soil saturated with KCl solution of a given concentration changed and increased according to the functions published in the results.

We have also developed a calculation model that can provide guidance for calculating the selective salinity of soil in laboratory conditions using the analytical application of conductometry. The model shows that by consistently changing the measurement frequencies, the selective salinity of a known soil can be determined by EC measurement under laboratory conditions. Because with solutions containing K^+ ions of the same concentration, the EC is saturated according to different functions depending on the measurement frequency.

As we advance our research on reference measurements, we would like to prove our thesis, by testing soils saturated with Ca^{2+} cations. If that test yields positive results, we will refine our model, under actual field conditions.

Acknowledgement

We would like to acknowledge the Hungarian University of Agriculture and Life Sciences, Institute of Technology for the financial support and the for the great interest work and concession of their areas allowing this work could be carried out. The authors would like to thank Zoltán Varga and Imre Czinkota for their intellectual suggestions when preparing the article.

References

[1] H. W. Griepentrog, 'Smart Crop Farming', *Landwirtschaftlicher Hochschultag*, *Landinfo Heft 3*, pp. 13-14, 2017

[2] H. Li *et al.*, 'Field-Scale Characterization of Spatio-Temporal Variability of Soil Salinity in Three Dimensions', *Remote Sensing*, Vol. 12, No. 24, Dec. 2020

[3] J. Horváth and B. Schmitz, 'Digitalisation in agriculture – From the perspective of a global agricultural machinery producer', *Hungarian Agricultural Engineering*, Vol. 36, pp. 63-68, 2019

[4] J. Popp *et al,* 'The prospects of precision agriculture', *International Journal of Engineering and Management Sciences*, Vol. 3, No. 1, Art. No. 1, Feb. 2018 (in Hungarian)

[5] J. V. Sinfield *et al.*, 'Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils', *Computers and Electronics in Agriculture*, Vol. 70, No. 1, pp. 1-18, Jan. 2010

[6] 'Agricultural Sensors Market'. https://www.reportsanddata.com/report-detail/agricultural-sensors-market (accessed Aug. 28, 2022)

[7] J. Hajdú, 'Quick soil testing and soil mapping', *Agronapló* 22, pp. 65-67, 2018 (in Hungarian)

[8] V. I. Adamchuk and R. A. V. Rossel, 'Development of On-the-Go Proximal Soil Sensor Systems', in *Proximal Soil Sensing*, R. A. Viscarra Rossel, A. B. McBratney, and B. Minasny, Eds. Dordrecht: Springer Netherlands, pp. 15-28, 2010

[9] R. A. Viscarra Rossel *et al.*, 'Chapter Five - Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time', in *Advances in Agronomy*, Vol. 113, D. L. Sparks, Ed. Academic Press, pp. 243-291, 2011

[10] S. Grunwald *et al.*, 'Fusion of Soil and Remote Sensing Data to Model Soil Properties', *Advances in Agronomy*, Vol. 131, D. L. Sparks, Ed. Academic Press pp. 1-109, 2015

[11] E. D. Lund, 'Soil Electrical Conductivity', in *Soil Science Step-by-Step Field Analysis*, S. Logsdon, D. Clay, D. Moore, T. Tsegaye, Eds. John Wiley & Sons, Ltd, 2008, pp. 137-146, 2008

[12] G. Galbács *et al.*, *Instrumental analytical chemistry exercises*, 3rd Edition. Szeged: JATEPress, 2015 (in Hungarian)

[13] D. L. Corwin and S. M. Lesch, 'Apparent soil electrical conductivity measurements in agriculture', *Computers and Electronics in Agriculture*, Vol. 46, No. 1-3, pp. 11-43, 2005

[14] J. D. Rhoades *et al.*, 'Soil Electrical Conductivity and Soil Salinity: New Formulations and Calibrations', *Soil Science Society of America Journal*, Vol. 53, No. 2, pp. 433-439, 1989

[15] J. D. Rhoades *et al.*, 'Geospatial measurements of soil electrical conductivity to assess soil salinity and diffuse salt loading from irrigation', *Geophysical Monograph-American Geophysical Union*, Vol. 108, pp. 197-216, 1999

[16] P. Stefanovits *et al.*, *Soil science*, Budapest: Mezőgazda, 2005 (in Hungarian)

[17] I. Waluyo *et al.*, 'The structure of water in the hydration shell of cations from x-ray Raman and small angle x-ray scattering measurements', *The Journal of chemical physics*, Vol. 134, No. 6, 2011

[18] H. Ohtaki and T. Radnai, 'Structure and dynamics of hydrated ions', *Chemical reviews*, Vol. 93, No. 3, pp. 1157-1204, 1993

[19] Y. Marcus, 'Effect of Ions on the Structure of Water: Structure Making and Breaking', *Chemical reviews*, Vol. 109, No. 3, pp. 1346-1370, 2009

[20] P. Jungwirth and D. J. Tobias, 'Specific Ion Effects at the Air/Water Interface', *Chemical reviews*, Vol. 106, No. 4, pp. 1259-1281, 2006

[21] S. Ghosal *et al.*, 'Electron Spectroscopy of Aqueous Solution Interfaces Reveals Surface Enhancement of Halides', *Science*, Vol. 307, No. 5709, pp. 563-566, 2005

[22] V. S. J. Craig and C. L. Henry, 'Specific Ion Effects at the Air–Water Interface: Experimental Studies', in *Specific Ion Effects*, WORLD SCIENTIFIC, pp. 191-214, 2009

[23] Dayo-Olagbende, G. O. and Ewulo, B. S, 'Ionic Mobility of Cations as Affected by Redox Status of Two Different Soil Textures', *Bulgarian Journal of Soil Science*, Vol. 6, No. 1, pp. 18-32, 2021

[24] P. Rengasamy, 'Processes involved in sodic behaviour. In 'Sodic soils. Distribution, properties, management, and environmental consequences'. (Eds M. E. Sumner, R. Naidu) pp. 35-50'. New York Press: New York, 1998

[25] P. Rengasamy, 'Soil Chemistry Factors Confounding Crop Salinity Tolerance—A Review', *Agronomy*, Vol. 6, No. 4, 2016

[26] J. Luo *et al.*, 'Distinctive Trend of Metal Binding Affinity via Hydration Shell Breakage in Nanoconfined Cavity', *The Journal of Physical Chemistry C*, Vol. 123, No. 23, pp. 14825-14833, 2019

[27] S. Hossain, 'Present Scenario of Global Salt Affected Soils, its Management and Importance of Salinity Research', *International Journal of Biological Sciences*, Vol. 1, No. 1, pp. 1-3, 2019

[28] B. Wicke *et al.*, 'The global technical and economic potential of bioenergy from salt-affected soils', *Energy & Environmental Science*, Vol. 4, No. 8, pp. 2669-2681, 2011

[29] F. I. Massoud and R. FAO, 'Salt affected soils at a global scale and concepts for control', Rome (Italy) FAO, 1981

[30] X. Pei *et al.*, 'Development and Performance Test for a New Type of Portable Soil EC Detector', *6th Computer and Computing Technologies in Agriculture (CCTA)*, Vol. AICT-392, No. Part I, p. 418, 2012

[31] M. R. Seifi *et al.*, 'Design and development of a portable soil electrical conductivity detector.', *Asian Journal of Agricultural Sciences*, Vol. 2, No. 4, pp. 168-173, 2010

[32] M. Li *et al.*, 'Development and performance test of a portable soil EC detector', *Applied Engineering in Agriculture*, Vol. 22, No. 2, pp. 301-307, 2006

[32] M. Trosin *et al.*, 'Measuring Soil Surface Roughness with the RealSense D435i', *Acta Polytechnica Hungarica*, Vol. 18, No. 6, pp. 141-155, 2021

[33] Md. S. I. Khan *et al.,* 'IoT and Wireless Sensor Networking-based Effluent Treatment Plant Monitoring System', *Acta Polytechnica Hungarica*, Vol. 18, No. 10, pp. 205-224, 2021

Modular Type of Learning Management System Services Formation based on Semantic Proximity

Saule Kumargazhanova, Yevgeniy Fedkin, Saule Smailova, Natalya Denissova, Saule Rakhmetullina, Yelena Blinayeva

D. Serikbayev East Kazakhstan Technical University
Faculty of Information Technology and Intelligent Systems
A. K. Protazanov Str. 69, 070004, Ust-Kamenogorsk, Kazakhstan e-mail:
{Saule.Kumargazhanova, Evgeny.Fedkin, SSmailova, NDenissova,
SRakhmetullina, EBlinaeva}@edu.ektu.kz

György Györök

Óbuda University, Alba Regia Technical Faculty, Budai út 45, H-8000
Székesfehérvár, Hungary e-mail: gyorok.gyorgy@amk.uni-obuda.hu

Abstract: Despite the many ready-made LMS solutions, there is a problem of creating LMS systems with a flexible architecture and meeting all user requirements. When creating such systems, it is difficult to take into account all the requirements for the functionality of the system. This paper proposes an approach to the formation of a service-oriented LMS based on modules. At the same time, depending on the functionality, the user can form the necessary functionality based on services, processes and objects. A cluster approach is used to form a finite set of modules.

Keywords: learning management system; online education; service-oriented approach; clustering; information technology; semantic proximity

Introduction

The use of modern information technologies in the educational process has led to the emergence of a special category of information systems – learning management systems (LMS - learning management system). This category provides a certain set of services that are focused on ensuring the implementation of the educational process at various levels [1, 2].

Over the past few years, the most interesting approach to software development has been based on user preferences and experience, the so-called user-oriented design [3-5]. This method is gaining more and more popularity among software developers, including those intended for training [6].

Recently, a new type of e-learning system has appeared, based on automatic recognition and prediction of user preferences, and self-adaptation to user requirements. This type of system is called adaptive learning management. Some researchers base training on the use of social networks and communication between trainees (students) and trainers (teachers) [7-9].

Some of current LMS systems are cloud-based and do not require a system administrator with experience in installing and maintaining software; however, there are also systems in which deep knowledge of programming languages such as PHP, JavaScript, knowledge of database management and administration, such as MySQL, Microsoft SQL Server is absolutely necessary [11].

Currently, the development and implementation of modern information and communication technologies in the educational process entails the transformation and search for new approaches to the implementation of LMS. One of these approaches is a service-oriented approach based on building a system from heterogeneous, loosely coupled parts, based on services as components with stationary interfaces and performing certain functions.

The study proposes a service-oriented architecture (SOA) approach that loosely connects various system components to reduce development time and costs [10]. In addition, a prototype of a training application with an API ecosystem is proposed, which takes into account the attractiveness and checks the functional and non-functional requirements for the system.

In this study, the authors determined which main categories of services are available in the LMS and proposed a reference model of LMS services. During the study, LMS services are divided into modules that implement similar functionality. On the basis of dedicated services, LMS modules of a specific educational institution are formed based on the requirements that are determined by this educational institution.

1 Materials and Methods

The methodology of this study consists in defining a set of services, forming a reference model, combining LMS services into modules based on the principles of semantic proximity and visualization of the results obtained.

1.1 Types of LMS Services

The functionality of LMS systems for a particular educational institution can solve various tasks depending on the features of the educational process in the educational institution. To solve problems, the LMS must provide a specific set of services that provide various aspects of the system.

The main LMS services can be classified into one of the following categories of services [12]:

1) Services for monitoring and controlling users. The services of this category are focused on fixing and evaluating the activity of various categories of users in various sections: user activity on a separate course and in the system as a whole, conducting various forms of control, evaluating various tasks performed by students on the course and forming the final grade for the course, etc.

2) Content monitoring and control services. Services of this category are focused on the support and control of methodological support of training courses based on a given set of requirements that are determined by the requirements of the educational process in a particular educational institution. These services can evaluate content formally (for example, the presence of certain educational components in a training course), and on the basis of expert evaluation.

3) Operational management and administration services. Services of this category are designed for operational management of work processes in the system. Services of this category can be divided into 2 subgroups:

- Services of operational management of the educational process. This subgroup of services is focused on the prompt response of problems that may arise during the educational process based on the control of certain parameters: information about the availability of courses for training, registration of students for courses, the occupancy of courses with various educational materials, etc.
- Hardware and software control services. This subgroup of services allows you to monitor and respond to problems that may arise in the operation of LMS software or hardware based on the collection and analysis of technical information about the activities of these components.

4) Analytics services. This category of services is aimed at carrying out analytical processing of various information contained in the LMS, as well as building various analytical reports and recommendations.

Based on the presented categories of LMS services and to implement the requirements for LMS systems, the authors propose a reference model of services for LMS systems.

1.2 Reference Model of LMS System Services

The reference model of LMS services is shown in Figure 1 [12].

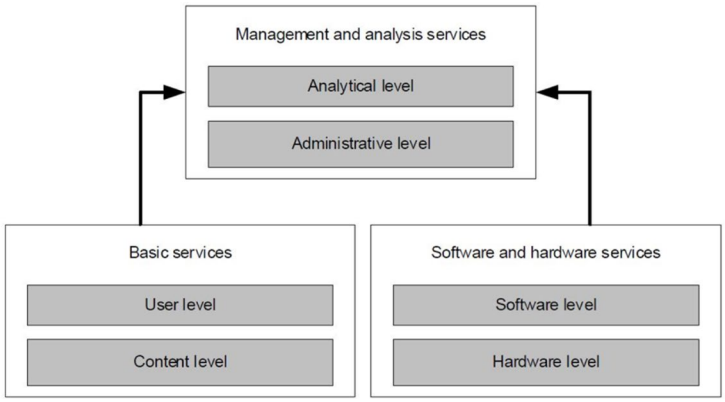


Figure 1
Reference model of LMS services

As can be seen from the diagram shown in Figure 1, LMS services are divided into 3 groups:

1) Hardware and software services. These services are represented by hardware and software components on the basis of which the LMS operates. These services are divided into the following levels:

- Hardware level. This level includes various types of hardware necessary for LMS operation (network equipment, telecommunication data transmission lines, servers, etc.)
- Software level. This level includes a variety of software that is involved in the operation of the LMS (operating systems, application servers, web services, Database Management Systems, etc.)

2) Basic services. These services are designed to provide the main function of LMS systems - training. These services are divided into the following levels:

- The level of content. This level includes services related to the support of working with various educational materials, conducting control activities, etc.
- User level. This level includes services related to the work of various categories of users in the LMS (user registration in the system, sharing access to various types of educational resources, logging user actions, etc.)

3) Management and analysis services. These services are designed to monitor the operation of the LMS system and conduct analysis based on data that accumulates during the operation of software and hardware services and basic services. These services are divided into the following levels:

- Administrative level. At this level, services aimed at operational management of the educational process are presented.
- Analytical level. At this level, services are provided for conducting analytical analysis based on data that was collected during the operation of the LMS.

The division of LMS services into the groups and levels listed above allows you to split services based on their functionality and combine them into functional modules that can be presented in LMS systems.

1.3 Combining LMS Services into Modules

Combining LMS services into modules will solve the following tasks [19]:

1) Determine the list of services that should be provided in the LMS, depending on the requirements of a particular educational institution.

2) Manage changes to LMS requirements. Changing the LMS requirements in this case can be solved in one of the following ways:

- by including and/or excluding modules from the LMS;
- inclusion of a new service in an existing module;
- creating a new module and including it in the LMS.

The following principles will be used to split the services:

1) The module is a specific set of services similar in their characteristics

2) The module has the possibility of expansion, which allows you to include new services in the existing module that are similar in their characteristics to the services available in the module

3) Extensibility of the module system. In case of new LMS services, the list can be revised by including new models and redistributing services between modules.

4) The presence of dependencies between modules. This principle assumes that any service of one module may depend on one or more services of other modules and, accordingly, the inclusion of some module in the LMS requires the inclusion of dependent modules.

1.4 Methodology of Combining LMS Services into Modules

As was mentioned in section 1.3, modules should be combined into services with similar characteristics. That is, the integration of services into modules should be carried out on the basis of their semantic proximity. In this study, the semantic proximity of services is determined by the following formula 1 [13, 14]:

$$SI(f_x, f_y) = \sum_{i=1}^{n} w_i SI_i \quad (1)$$

where, SI_i – is a measure of the *i*-connection between two services, w_i – is the weighting factor for the *i*-link.

To determine the semantic proximity of services, the following measures of connections and weight coefficients for them were used (Table 1) [13, 14]:

Table 1

Link measures and weighting factors

Designation	Link measure description	Weight coefficient
SI_{BP}	The ratio of the cardinal set of processes using both services to the cardinal number of the set of processes using at least one service	0,5
SI_{ER}	The ratio of the cardinal set of entities that are read in both services to the cardinal number of the set of services that are read by at least one of the services	0,27
SI_{EW}	The ratio of the cardinal set of entities that change in both services to the cardinal number of the set of services that change the essence of at least one of the services	0,23

To conduct a study based on the proposed model of combining LMS services into modules, it is necessary:

• define a list of LMS services;

• determine the list of processes that are carried out in the LMS;

• determine the list of objects that operate in the LMS;

• determine which services are involved in the implementation of a certain LMS process;

• determine which objects are used in the service to read their data;

• determine which objects are being created or modified in the service;

• based on the similarity of services, combine them into modules.

Combining services based on their semantic proximity can be done using the clustering method. To implement clustering, the FOREL algorithm is used since the number of clusters can be calculated during the operation and is a priori unknown [15, 16]. The scheme of operation of the clustering algorithm is shown in Figure 2.

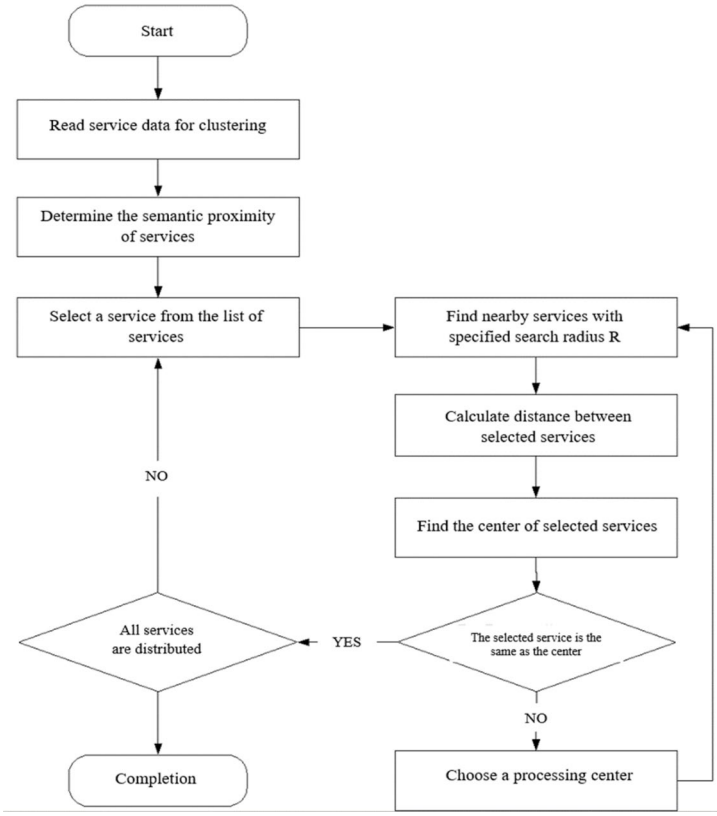


Figure 2
The scheme of the clustering algorithm Forel

2 Results

2.1 Approbation of the LMS Service Pooling Model

The study was conducted on the basis of the educational portal of D. Serikbayev East Kazakhstan Technical University [21], which has an LMS of its own design [17, 18].

Initially, the list of services that are available in the LMS on the educational portal and the dependence of the service on other services was determined (Table 2) [20].

Table 2

List of LMS services

Code	Service name	Depends on services	Service description
S01	Academic progress	S30, S32	student progress data is available. The report is equipped with filters by semester, a search system by discipline, by group, by the student's full name
S02	Unverified works and appeals	S23, S30	formation of a report on unverified works and submitted appeals on tests
S03	Task completion schedule	S07, S13, S23	uploading evaluated tasks and distance course tests
S04	Control of students and teaching staff in accordance with the schedule of classes	S33, S34	storing information about the educational process for a long time
S05	Active tests	S13, S30	contains data on students with active incomplete tests
S06	Export of academic progress	S23, S30, S31, S32	export of students' progress by control points of controls to the database of the educational portal
S07	Flexibility of setting up a system for evaluating students' work	S01, S03, S30, S31, S32	allows you to set up the evaluation of test tasks in accordance with the training system
S08	Proctoring	S18, S30	allows you to verify (confirm) the identity of the test taker, as well as observe his behaviour and what is happening on his computer screen to make sure that students are not cheating, does not resort to the help of other people, does not search for answers on the Internet
S09	Managing groups	S30, S32, S33	combining several discipline groups into one course
S10	Testing	S01, S30, S32	uploading tests for automatic control of students; control of the educational process: traditional, remote
S11	Teachers	S31, S33, S34	generating a report that allows you to view the activity of teachers in the distance learning system. Equipped with filters by semester, and inactivity of teaching staff by day, by full name of the teacher
S12	Student activity	S24, S30, S31, S32	the history of student visits allows you to see the log of user activity

S13	Test tasks	S23, S30, S31, S32	generating a report on uploaded tests for courses, equipped with filters by semester, only distant learning –generating a report by groups of students with the use of distance learning technologies by teachers
S14	Courses by type	S30, S31, S32, S33, S34	generating a report with quantitative indicators on loaded resources in the course section of the semester, form of study, school, department
S15	All courses	S10, S23, S30, S31, S32, S33	generating a report with data on all courses on filling courses with content in accordance with the parameters of technical expertise, equipped with filters by semester, schools, departments and a search system by Discipline, teacher, group
S16	Technical and methodological expertise service	S10, S23, S31, S33	conducting methodological expertise on course content and technical expertise on quantitative indicators of course resources
S17	Course academic progress	S01, S10, S23, S30, S31, S32, S33	viewing the progress of students, by group of a certain course
S18	File Storage		a file sharing service that provides the user of a distance learning system with a place for his files and round-the-clock access to them
S19	Video greeting	S18	allows you to upload a video greeting of the teacher with brief information about the course, requirements, wishes in any video format
S20	Course Forum	S24, S30, S32, S33	allows you to discuss course topics offline
S21	Chat	S30, S32, S33	allows you to discuss course topics online
S22	Private messages	S18, S30, S34	allows you to receive user messages
S23	Tasks	S18, S30, S31, S32, S33	allows you to upload a task for two-way communication
S24	User activity	S13, S21, S23, S30, S32, S33	based on the data of the log of user actions in the LMS, a diagram of activity in the E Monitoring system is built, analytical diagrams are formed on the activity of various categories of users in the LMS
S25	Examination session - Qualitative analysis	S01, S30, S31, S32	Formation of an analytical diagram based on the results of the examination session of students in the E Monitoring system

S26	Examination session - Academic progress	S01, S30, S31, S32	Formation of an analytical chart on the progress of students in the context of courses in the E Monitoring system
S27	Examination session - missed assignment	S01, S30, S31, S32	Formation of an analytical diagram on the missed assignments of students in the E monitoring system
S28	Video stream	S18	allows video conferencing
S29	Virtual boards	S18	a tool that allows you to visualize the training material
S30	Contingent accounting		Accounting of the contingent of students
S31	Training programs		taking into account the list of training modules to study in any educational program
S32	Individual training plan		Accounting of training modules for students to study according to a certain educational program
S33	Teacher's work plan		Accounting of training modules conducted by the teacher
S34	Personnel accounting	S33	Personnel accounting service

Next, a list of processes that go to the LMS and the services involved in each process is defined (Table 3).

Table 3

List of processes in the LMS and the services involved in them

Code	Process name	Services	Description of the process
P1	Formation of the course content	S03, S07, S10, S13, S14, S15, S18, S19, S23, S28, S31, S32	courses formation process: creating a course, uploading training materials, uploading training tasks, distributing points for each task, etc.
P2	Assessment of the course content	S03, S14, S15, S16, S23, S31	conducting an examination of the course content based on formal parameters and expert evaluation
P3	The learning process	S01, S02, S03, S05, S08, S09, S11, S12, S17, S20, S21, S22, S28, S29, S30, S31, S32, S33, S34	Conducting training courses
P4	Assessment of knowledge	S01, S02, S05, S06, S10, S17, S20, S21, S24, S25, S26, S27, S31, S32	Carrying out various forms of control measures within the framework of training courses
P5	System operation analysis	S01, S04, S05, S11, S12, S16, S24	Conducting an analysis of the system in order to identify problem areas in the work

At the next stage, the list of objects that are involved in the project is determined and for each object it is determined in which services the object data is read, and in which services the object is changed (Table 4).

Table 4
List of objects

Code	Object name	Reading Services	Change Services	Description of the object
O1	Student	S01, S02, S04, S05, S06, S08, S09, S10, S12, S13, S17, S18, S19, S20, S21, S22, S24, S25, S26, S27, S28, S29, S30, S32	S22, S30	The objects represent the student who is being trained in the courses
O2	Teacher	S02, S03, S04, S09, S10, S11, S13, S14, S15, S18, S20, S21, S22, S24, S28, S29, S33, S34	S22, S33, S34	the object represents a teacher who creates courses in the system and conducts classes
O3	Content	S03, S04, S07, S09, S10, S13, S14, S15, S16, S20, S21, S23, S24, S31	S07, S09, S18, S19, S20, S21, S23, S31	the object represents various types of materials that can be presented at courses (files, assignments, etc.)
O4	Academic performance	S30, S31, S32	S32	the object is a set of information about progress on various types of tasks in courses
O5	User activity	S02, S04, S05, S10, S11, S12, S20, S21, S23, S24, S33	S08, S10, S20, S21, S23, S28, S29, S33	the object is a log of the user's activity in the system and the artifacts generated by it
O6	Curriculum	S01, S04, S14, S15, S16, S17, S25, S26, S27, S31, S32, S33	S31, S32	the object represents a list of disciplines to study
O7	File storage	S18, S30, S33	S18	File storage

At the final stage, a list of software modules has been determined into which the services presented in Table 2 will be combined. To do this, a list of predefined modules was first defined (Table 5):

Table 5
Predefined modules

Code	Module name	Services
M01	Personnel accounting module	S34
M02	Contingent accounting module	S30
M03	Module of training programs	S31
M04	Curriculum Module	S32, S33

The remaining unallocated modules are formed based on formula 1. Semantic similarity of services is determined for the distribution. Figure 3 shows a matrix of semantic proximity of services.

	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	S29
S01	1.00	0.63	0.33	0.50	0.82	0.53	0.00	0.30	0.23	0.18	0.56	0.65	0.30	0.30	0.30	0.45	0.83	0.07	0.14	0.39	0.39	0.26	0.00	0.62	0.67	0.67	0.67	0.22	0.26
S02	0.63	1.00	0.42	0.39	0.74	0.57	0.00	0.34	0.39	0.37	0.58	0.58	0.37	0.28	0.28	0.23	0.80	0.14	0.09	0.70	0.70	0.43	0.07	0.60	0.55	0.55	0.55	0.35	0.43
S03	0.33	0.42	1.00	0.34	0.33	0.23	0.30	0.17	0.35	0.26	0.45	0.36	0.58	0.74	0.74	0.45	0.36	0.23	0.17	0.26	0.26	0.26	0.42	0.37	0.23	0.23	0.23	0.42	0.26
S04	0.50	0.39	0.34	1.00	0.50	0.28	0.05	0.05	0.16	0.22	0.59	0.59	0.39	0.39	0.39	0.59	0.34	0.09	0.05	0.22	0.22	0.11	0.11	0.70	0.34	0.34	0.34	0.11	0.11
S05	0.82	0.74	0.33	0.50	1.00	0.53	0.00	0.30	0.23	0.26	0.65	0.83	0.30	0.23	0.23	0.36	0.65	0.07	0.14	0.47	0.47	0.26	0.09	0.70	0.49	0.49	0.49	0.22	0.26
S06	0.53	0.57	0.23	0.28	0.53	1.00	0.00	0.27	0.09	0.32	0.23	0.37	0.32	0.23	0.23	0.23	0.62	0.09	0.27	0.32	0.32	0.14	0.00	0.55	0.87	0.87	0.87	0.14	0.14
S07	0.00	0.00	0.30	0.05	0.00	0.00	1.00	0.00	0.32	0.32	0.00	0.00	0.59	0.34	0.34	0.14	0.00	0.62	0.73	0.18	0.18	0.00	0.50	0.07	0.00	0.00	0.00	0.25	0.00
S08	0.30	0.34	0.17	0.05	0.30	0.27	0.00	1.00	0.59	0.30	0.25	0.39	0.09	0.00	0.00	0.00	0.39	0.09	0.27	0.43	0.43	0.64	0.12	0.07	0.14	0.14	0.14	0.62	0.87
S09	0.23	0.39	0.35	0.16	0.23	0.09	0.32	0.59	1.00	0.20	0.32	0.32	0.27	0.14	0.14	0.07	0.32	0.25	0.32	0.57	0.57	0.68	0.18	0.20	0.07	0.07	0.07	0.43	0.68
S10	0.18	0.37	0.26	0.22	0.26	0.32	0.32	0.30	0.20	1.00	0.14	0.14	0.45	0.27	0.27	0.05	0.22	0.36	0.32	0.55	0.55	0.14	0.42	0.44	0.30	0.30	0.30	0.53	0.37
S11	0.56	0.58	0.45	0.59	0.65	0.23	0.00	0.25	0.32	0.14	1.00	0.82	0.30	0.30	0.30	0.40	0.40	0.07	0.00	0.30	0.30	0.34	0.09	0.53	0.23	0.23	0.23	0.26	0.34
S12	0.65	0.58	0.36	0.59	0.83	0.37	0.00	0.39	0.32	0.14	0.82	1.00	0.30	0.23	0.23	0.40	0.49	0.07	0.14	0.30	0.30	0.34	0.09	0.53	0.32	0.32	0.32	0.26	0.34
S13	0.30	0.37	0.58	0.39	0.30	0.32	0.59	0.09	0.27	0.45	0.30	0.30	1.00	0.62	0.62	0.30	0.30	0.64	0.59	0.20	0.20	0.18	0.32	0.43	0.30	0.30	0.30	0.43	0.18
S14	0.30	0.28	0.74	0.39	0.23	0.23	0.34	0.00	0.14	0.27	0.30	0.23	0.62	1.00	1.00	0.58	0.30	0.30	0.25	0.11	0.11	0.07	0.57	0.34	0.30	0.30	0.30	0.23	0.07
S15	0.30	0.28	0.74	0.39	0.23	0.23	0.34	0.00	0.14	0.27	0.30	0.23	0.62	1.00	1.00	0.58	0.30	0.30	0.25	0.11	0.11	0.07	0.57	0.34	0.30	0.30	0.30	0.23	0.07
S16	0.45	0.23	0.45	0.59	0.36	0.23	0.14	0.00	0.07	0.05	0.40	0.40	0.30	0.58	0.58	1.00	0.32	0.00	0.00	0.05	0.05	0.00	0.26	0.45	0.32	0.32	0.32	0.00	0.00
S17	0.83	0.80	0.36	0.34	0.65	0.62	0.00	0.39	0.32	0.22	0.40	0.49	0.30	0.30	0.30	0.32	1.00	0.07	0.14	0.55	0.55	0.34	0.00	0.45	0.75	0.75	0.75	0.26	0.34
S18	0.07	0.14	0.23	0.09	0.07	0.09	0.62	0.09	0.25	0.36	0.07	0.07	0.64	0.30	0.30	0.00	0.07	1.00	0.71	0.18	0.18	0.18	0.33	0.11	0.07	0.07	0.07	0.43	0.18
S19	0.14	0.09	0.17	0.05	0.14	0.27	0.73	0.27	0.32	0.32	0.00	0.14	0.59	0.25	0.25	0.00	0.14	0.71	1.00	0.18	0.18	0.14	0.37	0.07	0.14	0.14	0.14	0.39	0.14
S20	0.39	0.70	0.26	0.22	0.47	0.32	0.18	0.43	0.57	0.55	0.30	0.30	0.20	0.11	0.11	0.05	0.55	0.18	0.18	1.00	1.00	0.39	0.37	0.44	0.30	0.30	0.30	0.42	0.50
S21	0.39	0.70	0.26	0.22	0.47	0.32	0.18	0.43	0.57	0.55	0.30	0.30	0.20	0.11	0.11	0.05	0.55	0.18	0.18	1.00	1.00	0.39	0.37	0.44	0.30	0.30	0.30	0.42	0.50
S22	0.26	0.43	0.26	0.11	0.26	0.14	0.00	0.64	0.68	0.14	0.34	0.34	0.18	0.07	0.07	0.00	0.34	0.18	0.14	0.39	0.39	1.00	0.00	0.14	0.09	0.09	0.09	0.52	0.77
S23	0.00	0.07	0.42	0.11	0.09	0.00	0.50	0.12	0.18	0.42	0.09	0.09	0.32	0.57	0.57	0.26	0.00	0.33	0.37	0.37	0.37	0.00	1.00	0.14	0.00	0.00	0.00	0.28	0.12
S24	0.62	0.60	0.37	0.70	0.70	0.55	0.07	0.07	0.20	0.44	0.53	0.53	0.43	0.34	0.34	0.45	0.45	0.11	0.07	0.44	0.44	0.14	0.14	1.00	0.53	0.53	0.53	0.14	0.14
S25	0.67	0.55	0.23	0.34	0.49	0.87	0.00	0.14	0.07	0.30	0.23	0.32	0.30	0.30	0.30	0.32	0.75	0.07	0.14	0.30	0.30	0.09	0.00	0.53	1.00	1.00	1.00	0.09	0.09
S26	0.67	0.55	0.23	0.34	0.49	0.87	0.00	0.14	0.07	0.30	0.23	0.32	0.30	0.30	0.30	0.32	0.75	0.07	0.14	0.30	0.30	0.09	0.00	0.53	1.00	1.00	1.00	0.09	0.09
S27	0.67	0.55	0.23	0.34	0.49	0.87	0.00	0.14	0.07	0.30	0.23	0.32	0.30	0.30	0.30	0.32	0.75	0.07	0.14	0.30	0.30	0.09	0.00	0.53	1.00	1.00	1.00	0.09	0.09
S28	0.22	0.35	0.42	0.11	0.22	0.14	0.25	0.62	0.43	0.53	0.26	0.26	0.43	0.23	0.23	0.00	0.26	0.43	0.39	0.42	0.42	0.52	0.28	0.14	0.09	0.09	0.09	1.00	0.75
S29	0.26	0.43	0.26	0.11	0.26	0.14	0.00	0.87	0.68	0.37	0.34	0.34	0.18	0.07	0.07	0.00	0.34	0.18	0.14	0.50	0.50	0.77	0.12	0.14	0.09	0.09	0.09	0.75	1.00

Figure 3

Matrix of semantic proximity of services

After determining the semantic proximity matrix, a computational experiment was carried out with the following search radii:

- *R= 0.5* – this radius corresponds to half of the maximum distance between services (Table 6)

Table 6

Modules obtained by clustering with a radius of 0.5

Code	Module name	Services
M05	The module of accounting of progress	S01, S06, S17, S25, S26, S27
M06	User Activity Module	S02, S05, S11, S12, S24
M07	Module of training tasks	S23
M08	Testing Module	S10
M09	Online communication module	S08, S09, S22, S28, S29
M10	Course content Control module	S03, S13, S14, S15
M11	Communication on the course module	S20, S21
M12	Course Examination Module	S16
M13	The module for monitoring the conduct of teaching staff classes	S04
M14	Course Setup Module	S07, S18, S19

- *R= 0.34* – this radius corresponds to the average value of all elements of the semantic proximity matrix (Table 7)

Table 7

Modules obtained by clustering with a radius of 0.34

Code	Module name	Services
M05	User Activity Module	S04, S05, S24
M06	Task Module	S23
M07	Academic Progress Monitoring Module	S01, S06, S17, S25, S26, S27
M08	Module of Test Tasks	S13
M09	Proctoring Module	S08
M10	Testing Module	S10
M11	Examination Module	S16
M12	Visit Control Module	S11, S12
M13	Course Occupancy Control Module	S03, S14, S15
M14	Video Stream Module	S28
M15	Communication Module	S09, S22, S29
M16	Task Control Module	S02, S20, S21
M17	Module for Configuring Data Storage for the Course	S07, S18, S19

As a result, 14 modules were obtained for a search radius of 0.5 and 17 modules for a search radius of 0.34 for the initial list of services. The graphical distribution of services by modules for different search radii is shown in Fig. 4.

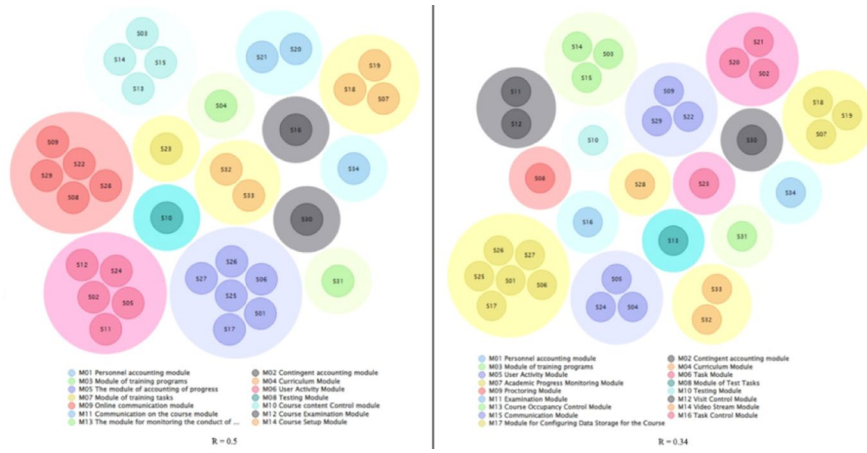


Figure 4

Distribution of LMS services by modules

After distributing the services by modules, you can determine the dependencies between the received modules based on the dependencies between the services.

For this study, the dependencies between the obtained modules are shown in Figure 5.

Figure 5
Distribution of LMS services by modules

The results of the experiment showed:

1) services similar in semantics were included in one cluster (module), for example, services S03, S13, S14, S15 were included in the module "Course content control module" at $R = 0.5$, at $R = 0.34$ services S03, S14, S15 were included in the module "Course content control module";

2) loosely coupled functions, regardless of the radius used to search for neighboring elements, are allocated to a separate cluster (service), for example, service S10;

3) the distribution of services with a high degree of connectivity depends on the clustering parameters, which makes it possible to identify these services in subsequent attempts, varying the radius of the search for neighbouring elements.

3 Software Implementation

To solve the above tasks, the authors have designed an information system database. The logical scheme of which is shown in Figure 6.

The developed database architecture of this study was implemented in the Mircosoft SQL Server 2019 database management system in the form of a physical LmsModules database.

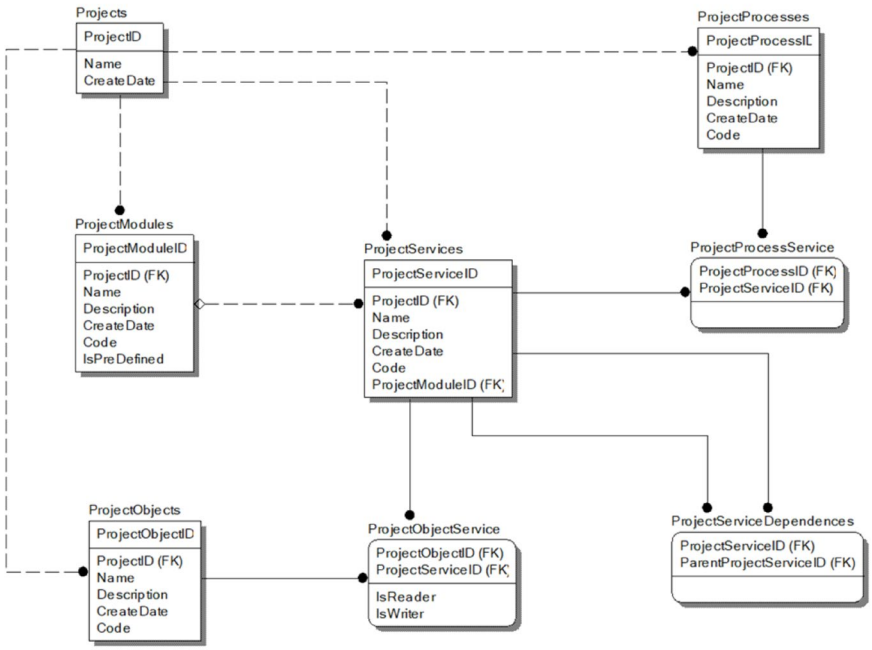


Figure 6
Logic diagram of the LMS modular construction database

At the final stage of this part of the study, a web application has been developed to work with the specified database. The architecture of the web application is shown in Figure 7. The web application is developed on the platform ASP.NET and it is divided into several levels:

1) The "DB Manager" component is located at the lower level. This component is responsible for interacting with the database described above in this section.

2) At the next level there is a "Module for working with database elements". This module contains components that allow you to extract data from our database and edit them.

3) The Forel Clusterer component implements mechanisms for distributing services across modules using the Forel clustering algorithm. To determine clusters, an assessment of the semantic proximity of services is used, according to the model described in Section 1.

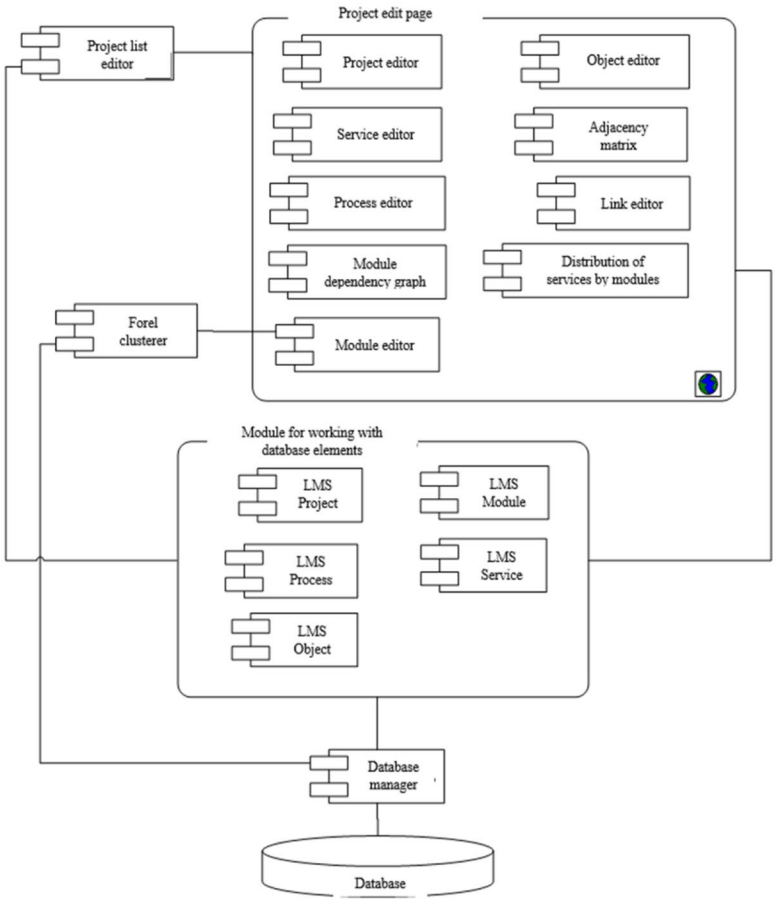


Figure 7
Web Application architecture

3) The "Project List Editor" page. On this page there is an option to create a new project, delete a project or copy an LMS project (Fig. 8).

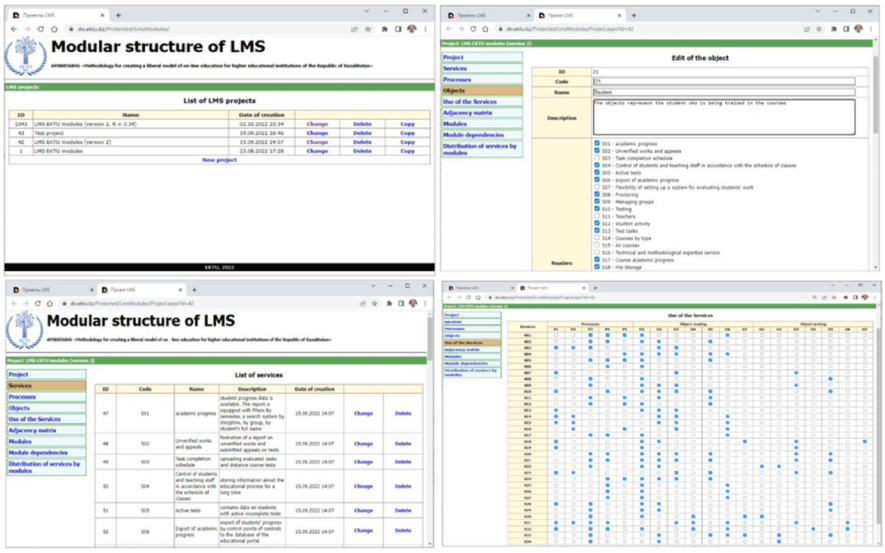


Figure 8
Web application appearance

4) The project editing page. On this page, the main work is done to create an LMS project (Fig. 8). Let's look at working with this page in more detail. This page contains the following elements:

- Project editor. This component of the page displays the name of the project, and also in this part we can adjust the name of the project.
- Service editor. This component of the page contains a list of services for the project and it is possible to add, delete or change data on the service.
- Process editor. This component of the page contains a list of processes that go on in the project and it is possible to add, delete or change process data
- Object editor. This component of the page contains a list of objects that are being worked with in the LMS project, and it is also possible to create, delete or modify data on objects.
- Link editor. This component of the page presents the possibility of setting the use of services in processes and the use of objects for reading and writing in services.
- Adjacency matrix. This component of the page displays a matrix of adjacency of services, which is based on the methodology presented in Section 1.
- Module editor. This component contains a list of modules that include project services. In this part, you can create, delete or modify a module. The page also presents the possibility of generating modules. When generating modules, all

previous modules are deleted, except the predefined ones. To generate modules, you can specify the cluster radius for Forel clustering.

• A graph of module dependencies. This component of the page displays a dependency graph between modules based on dependencies between services.

• Distribution of services by modules. This component of the page presents a graphical distribution of services according to the created models.

The developed web application allows you to form LMS modules based on the proposed clustering model.

The resulting software modules will increase the efficiency of their support and expansion, and also, based on the dependencies between the modules, it will allow you to predict how the expansion of functionality may affect other modules.

Conclusions

Within the framework of this study, a methodology for combining LMS services into modules based on semantic proximity is proposed. The integration of services is based on the processes in which the services are involved and on the objects with which the services work. As a clustering method, the Forel method was used, which allows clustering without a predefined number of clusters.

Based on the proposed methodology, a web application has been developed that allows the user to develop a unique LMS configuration based on its own preferences or requirements.

Acknowledgement

This work was supported by the Ministry of Education and Science of the Republic of Kazakhstan under the grant financing program for the 2020-2022 years by the program 08856846 “Methodology for creating a liberal model of Online education for higher education institutions of the Republic of Kazakhstan”.

References

[1] Walker, D., Lindner, J., Murphrey, T., Dooley, K. Learning management system usage: Perspectives from university instructors, Quarterly Rev. Dist. Ed., 2016, 17 (2), 41-50

[2] DeRouin, R., Fritzsche, B., Salas, E. E-learning in organizations, Journal of Management, 2005, 31 (6), 920-940

[3] Wang, T. Y., Wang, C. H. E-Learning Platform of STEAM Aesthetic Course Materials Based on User Experience. 1st International Cognitive Cities Conference (IC3) 2018, 123-128

[4] Wong, S. K., Nguyen, T. T., Chang, E., Jayaratna, N. Usability metrics for E-learning Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2003, 235-252

[5] Zaharias, P., Pappas, C. Quality management of learning management systems: A user experience perspective Current Issues in Emerging eLearning, 2019, Vol. 3: Iss. 1, Article 5, 60-83

[6] Kraleva, R. S. Designing an Interface for a Mobile Application Based on Children's Opinion International Journal of Interactive Mobile Technologies (iJIM), 2017, 11(1), 53-70

[7] Lin, J.-W., Lin, H.-C. K. User acceptance in a computer-supported collaborative learning (CSCL) environment with social network awareness (SNA) support, Australasian Journal of Educational Technology, 2019, 35(1), 100-115

[8] Narayan, V., Herrington, J., Cochrane, T. Design principles for heutagogical learning: Implementing student-determined learning with mobile and social media tools, Australasian Journal of Educational Technology, 2019, 35(3), 86-101

[9] Valova, I., Marinov, M. Facebook as a Tool Aiding University Education-Whether it is Possible and Useful, TEM Journal, 2019, Vol. 8(2), 670-676

[10] Sun, G., Huang, Z., Yue, L. Building API Ecosystem in Blended Learning by Service Oriened Architecture, 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI), 2022, 673-678

[11] WCET: WICHE Cooperative for Educational Technologies. Available online: http://wcet.wiche.edu (accessed on 10.05.2023)

[12] Fedkin, Y., Smailova, S., Zuev, V., Kumargazhanova, S., Denissova, N., Aitmukhanbetova, E. Classification of lms control and monitoring services, Collection of materials of the VI International Scientific and Practical Conference "Computer Science and Applied Mathematics", 2021 September 29 - October 2, Almaty, Kazakhstan, 232-244

[13] Mokerov, V. Development of methods and models for designing services of process-oriented information systems, Thesis for the degree of Doctor of Philosophy (PhD): 6D070300, Ust'-Kamenogorsk, 2014

[14] Mokerov, V. An approach to service-oriented information systems architecture development based on semantic closure measure, 2014 Federated Conference on Computer Science and Information Systems, FedCSIS 2014, 2014, September 7 – 10, WARSAW, Poland, 269-272

[15] Likhtina, E. V. COMPARISON OF CLUSTERING ALGORITHMS, Actual problems of aviation and cosmonautics, 2018, Vol. 2, No. 14, 73-74

[16] Gorshkov, S. N., Kalashnikova, V. A., Solomov, D. V. The study of the cluster analysis model of the content of Web resources, Innovations in science, 2019, No. 1 (89), 7-10

[17] Fedkin, E., Kumargazhanova, S., Smailova, S., Denissova, N., György Györök. Considering the functioning of an e-learning system, based on a model for assessing the performance and reliability of the system, Acta Polytechnica Hungarica, 2022, Vol. 19, No. 2, 93-112

[18] Mutanov, G., Shakarimova, A. Educational portal of the university: Theory and practice; Publisher: East-Kazakhstan State Techinical University, Kazakhstan, Ust-Kamenogorsk, 2006, p. 352

[19] Maakot, A. METHODOLOGY FOR DESIGNING THE MODULAR STRUCTURE OF THE INFORMATION SYSTEM, Izvestiya SFedU. Engineering Sciences, 2017, No. 3 (188), 42-51

[20] Zuev, V., Kakisheva. L., Denissova, N., Kumargazhanova, S., Smailova, S. Development of a Set of Requirements for the Hardware and Software of LMS Services of the University, International Journal of Emerging Technologies in Learning (iJET), 2021, Vol. 16, No. 21, 210-218

[21] Educational portal of D. Serikbayev EKTU. Available online: https://www.do.ektu.kz (accessed on 10.10.2022)

Design of Dynamically Balanced Gait for the Biped Robot While Crossing the Ditch

Moh Shahid Khan*, Ravi Kumar Mandava

Mechanical Engineering Department, Maulana Azad National Institute of Technology, Bhopal, Madhya Pradesh, India 462003;
E-mail: 203116026@stu.manit.ac.in; ravikumar1013@manit.ac.in
*Corresponding author

Abstract: The current research work aims to generate the dynamically balanced gait for the 16-DOF biped robot while crossing the ditch by using the concept of the zero moment point (ZMP). Initially, forward kinematics was established to obtain the position and orientation of the biped robot while crossing the ditch. The various joint angles of the biped robot were estimated by deriving the inverse kinematics. Further, the dynamics of the biped robot was obtained using the Lagrange-Euler formulation. A cubic polynomial equation was assigned for the smooth motion of foot and wrist trajectories in the sagittal plane and hip trajectory in the horizontal plane. The obtained cubic polynomial trajectory for the foot was compared with the second-order and fifth-order polynomial trajectories in terms of dynamic balance margin (DBM). A simulation study was conducted to verify the dynamically balanced gait while crossing the ditch. Finally, the generated gait angles were tested on a real 16-DOF biped robot. It has been found that the generated gait is more dynamically balanced while crossing the ditch.

Keywords: Biped robot; gait generation; ditch crossing; DBM; ZMP; Lagrange-Euler formulation

1 Introduction

Bipedal robots have many advantages over other wheeled robots due to their similar characteristics to human beings. That similarity creates complexity for building a biped robot and enabling it to perform human actions due to the very complex and multi-degrees of freedom (DOF) mechanism. The most crucial and significant requirement for developing biped robots is the ability to move across various terrains. Around the world, many researchers are working on the dynamically balanced gait generation of the biped robot on various uneven terrains. The gait cycle consists of two phases that is, the single support phase (SSP) and the double support phase (DSP). The dynamic stability of the biped robot during the gait cycle can be obtained from the concept of the zero moment point (ZMP) [1]. Chow and Jacobson [2] applied the Lyapunov function, the linear feedback principle, and an

on-off perturbation to generate the gait on a flat surface. Townsend and Tsai [3] generated a variety of gaits on both SSP and DSP by using various initial parameters. In addition, Katoh and Mori [4], [5] presented a control method for a 4-DOF biped robot by using Van Der Pol's equation and obtained stable limit cycles. Mita et al. [6] generated the Chiba Walker 1 (CW-1) mechanism, which requires one second to complete each step with a step length of less than 20 cm. Further, Hürmüzlü and Moskowitz [7] developed a mathematical model based on LIPM (linear inverted pendulum model) to address the impact of periodic force on the stability of biped locomotion. Takanishi et al. [8] stabilized the gait patterns while walking under the influence of a known external unbalanced force produced by a DD (direct drive) motor for the WL-12R biped robot. Then, the researchers proposed a control method that recognizes the geometry of the terrain, which allowed the biped robot WL-12RIII to walk on stairs with a 0.1-meter step height and $\pm 10^{\circ}$ inclined trapezoidal terrain [9]. Moreover, Zheng [10] discussed an autonomous gait generation for biped robots to walk on complex terrains with the help of a central pattern generator (CPG) and the concept of neural network (NN). Later on, Thomas Miller [11] proposed a hierarchical control for a 10-axis biped model using a PID controller and a cerebellar model arithmetic computer (CMAC) NN learning system. Also, Arakawa and Fukuda [12] established an ideal gait for a 13-DOF biped robot by reducing energy consumption by using the concept of ZMP and genetic algorithm (GA). To generate the stable gait for the 6-link biped robot, Magdalena and Monasterio-Huelin [13] developed a learning mechanism by using GA and a fuzzy logic controller (FLC). GA helps to modify the fuzzy rules and their functions relevant to the human information database. Abba & Chaillet [14], established bidirectional dynamic modelling by using the features of the epicyclic gear train, and computed torque control (CTC). In [15], the authors used a V-HRP (Virtual Humanoid Robot Platform) simulator to verify the robot simulation virtually before going to test in a real-time environment. The said robot consists of 26-DOF and a height of 540 mm equipped with a CCD camera, foot sensors, a posture sensor, and a USB. Chestnutt et al. [16] proposed an architecture for an H7 humanoid robot to walk on uneven ground along with obstacles. In addition, Sabourin and Bruneau [17] discussed a CMAC-NN-based control strategy for fast walking of a virtual under-actuated biped robot. After learning, the NN first generates the passive and active gaits of the biped robot. Puga et al. [18] obtained a distributed control scheme that consists of a fractional-order PID controller optimized by a GA. The control parameters were tested on a 22-DOF small-sized humanoid robot. Further, Ghorbani et al. [19] developed a control scheme using a general regression NN (GRNN) feedback control and a PID feedback controller with Lyapunov exponents to stabilize the LIPB-based biped model. A Genetic algorithm is used to optimize the GRNN controller to reduce energy consumption and create a closed-loop learning controller.

Vundavilli and Pratihar [20] suggested an analytical method for a 7-DOF biped robot with the help of the inverse dynamics trained neural network (NN) to generate the gait on staircases, sloping surfaces, and ditches. Further in [21], [22], they

discussed the optimal online gait generation of the biped robot using NN and FL-based gait planners which was trained by GA. In addition, Fattah and Fakhari [23] established a trajectory planning algorithm for a seven-link planar biped robot with variable step lengths on level ground along with ditches. Sudheer et al. [24] suggested a framework for an eight-link biped robot to solve the kinematics and dynamics based on the ZMP constraint and optimized using a simulated annealing approach. A cycloidal trajectory was assigned for the swing foot and the modified Cartesian cycloid trajectory for the hip. Kalamian and Farrokhi [25] proposed a nonlinear model predictive control (NMPC) algorithm to generate the gait while crossing the obstacle at a speed of 1 m/s. Moreover, Lathan et al. [26] suggested an analytical method of a 7-DOF biped robot to generate a dynamically balanced gait while crossing and stepping over the obstacle. R. Kumar et al. [27] developed the neural network (NN) based gait planners for a biped robot for crossing obstacles. The weights of the NN algorithm were optimized by GA and differential evolution (DE) algorithms. Later, Mandava and Vundavilli [28]–[30] developed a framework for generating whole-body motions of an 18-DOF humanoid robot using the concept of inverse kinematics. Bai et al. [31] discussed a dynamically balanced gait generation of the humanoid robot while crossing over obstacles with the help of a pulsing type joint. The said humanoid robot is capable of squatting and displacing heavy objects from one place to another. Furthermore, Anh and Huan [32] established a method for optimizing the gait generation of a 10-DOF biped robot. The modified Jaya algorithm was used to optimize the distance between the ZMP and the foot center then said approach was tested on HUBOT-4. Also, Tsuru et al. [33] proposed a structure of an autonomous humanoid robot for finding and holding objects in unknown surroundings. The proposed methodology was experimentally applied to HRP2-KAI and validated its efficiency. Yang et al. [34] discussed an NN estimator with an incremental learning mechanism while obtaining a new online walking controller for biped robots. The proposed method controls the ZMP stability while compensating for the yaw moment. Later on, researchers developed various methodologies to generate the gait for avoiding obstacles, stepping over them, and crossing ditches. Kashyap et al. [35] developed an Improved Modified Chaotic Invasive Weed Optimization (IMCIWO) algorithm to navigate the humanoid robot and avoid obstacles. The authors conducted a simulation on the NAO humanoid robot in WEBOT software. In [36], established a model predictive controller (MPC) for NAO humanoid robot while walking on uneven terrain, in an unknown environment and stepping over obstacles. Janardhan and Kumar [37] proposed a multibody dynamics framework for gait generation of the 5-DOF biped robot while crossing the ditches. The width of the ditch is greater than the length of the leg. Gupta and Dutta [38] developed a trajectory and navigation planner for a 12-DOF biped robot while walking on uneven terrain along with obstacles.

Apart from the above discussion, some researchers have developed various techniques for generating a dynamically balanced gait for the biped robot as it crosses the ditch. To generate the systematic gait a cubic polynomial trajectory was assigned for the foot in the sagittal plane and the hip in the horizontal plane.

In addition, an inverse kinematics approach is established for obtaining the various joint angles of the biped robot along with foot trajectory. Moreover, the balancing of the biped robot is calculated using the concept of ZMP and measured in terms of DBM. Further, the dynamics of the biped robot is obtained by using the Lagrange-Euler formulation which is useful to determine the torque required for each joint of the robot. Additionally, it has been noted that only a few researchers have developed real-time biped robots for verifying the gait while crossing a ditch. In the present research work, the authors have developed a real biped robot titled AZAD-16, which consists of 16-DOF for verifying the simulations in a real-time environment.

The rest of the article is organized as follows: Section 2 explains the mathematical modelling of the biped robot, which includes the physical dimensions of AZAD-16, trajectory generation for the foot, hip, and arm, forward and inverse kinematics, DBM as fundamental stability criteria, and the Lagrange-Euler (LE) algorithm for estimating the dynamics. Further, the results and discussions of the current research work are discussed in Section 3. Finally, Section 4 provides the conclusions of the proposed research work.

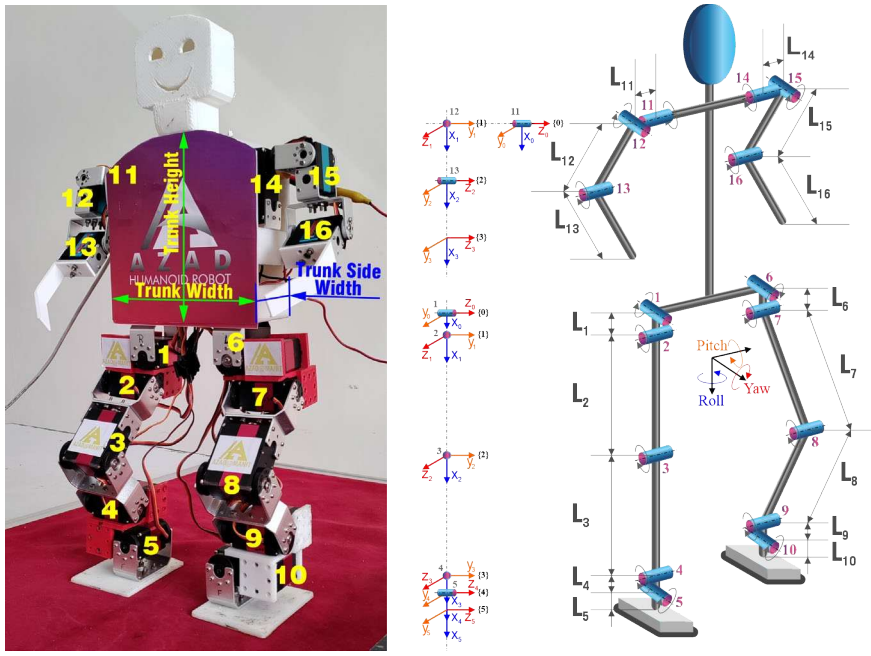
2 Mathematical Modelling

AZAD-16 is a tiny biped robot that was built in our laboratory which is shown in Figure 1(a). The trunk and various brackets for holding the servo motors were made by 3D printing. All the joints of the 16-DOF biped robot consist of rotatory joints which are shown in Figure 1(b). The weight and height of the robot is 5 kg and 480 mm. Each leg of the biped robot consists of 5-DOF mobility which contributes 2-DOF for hip joint, 2-DOF for the ankle joint and 1-DOF for the knee joint. However, each arm has 3-DOF mobility, which contributes 2-DOF for the shoulder and 1-DOF for the elbow joint. The joints of the legs are attached with 60 kg-cm rated servo motors and the joints of the arms are attached with 35 kg-cm rated servo motors. Various input parameters such as power rating, type of actuation, link length and mass related to the AZAD-16 biped robot are shown in Table 1.

Table 1
Various input parameters of the "AZAD-16"

AZAD-16	Joints Name	Power Rating (Kg-cm)	Type of Actuation	Link's Name	Link Length (mm)	Link Mass (Kg)
Right Leg	Hip	60	Yaw	-	35.12	0.15
		60	Pitch	Thigh	93.87	0.25
	Knee	60	Pitch	Shank	67.00	0.25
	Ankle	60	Pitch	-	35.24	0.15
		60	Yaw	-	42.50	0.15

Left Leg	Hip	60	Yaw	-	35.12	0.15
		60	Pitch	Thigh	93.87	0.25
	Knee	60	Pitch	Shank	67.00	0.25
	Ankle	60	Pitch	-	35.24	0.15
		60	Yaw	-	42.50	0.15
Right Arm	Shoulder	60	Pitch	-	33.00	0.25
		35	Yaw	Arm	93.76	0.1
	Elbow	35	Pitch	Forearm	93.76	0.1
Left Arm	Shoulder	60	Pitch	-	33.00	0.25
		35	Yaw	Arm	93.76	0.1
	Elbow	35	Pitch	Forearm	93.76	0.1
Trunk	-	-	-	Height	175.00	0.3
				Front Width	165.00	
				Side Width	70.00	



(a) (b)

Figure 1
Schematic diagram showing the (a) real biped robot that is AZAD-16 and (b) kinematic model displaying the positioning and coordinate frames for various joints

2.1 Trajectory Generation for Foot, Hip, and Arm

The foot, hip, and arm of the robot AZAD-16 follow a polynomial trajectory while crossing the ditch. The terms z and x have been considered the height and length of the polynomial at a particular time interval. Where, $\mu_0, \mu_1, \mu_2 \ldots \ldots \mu_n$ have been taken as coefficients of the polynomial equation, respectively. The initial and final positions of the foot have been considered x_1, x_2 and x_3. In the present research problem, the foot trajectory of the robot is considered quadratic, cubic and fifth ordered polynomial equations and compared the results in terms of dynamic balance margin. Further, hip trajectory in horizontal plane and wrist trajectory in sagittal plane have been assigned as cubic polynomial trajectory. The distances between the trunk and wrist end at initial and final instances of wrist trajectory are x_{01} and x_{03}, respectively. The height of the hip joint and width of the ditch are represented as H and d_w. The boundary conditions for generating various polynomial trajectories for the foot, hip and wrist are shown in Table 2.

Table 2
Boundary conditions for different trajectories of wrist, hip and swing foot in various planes

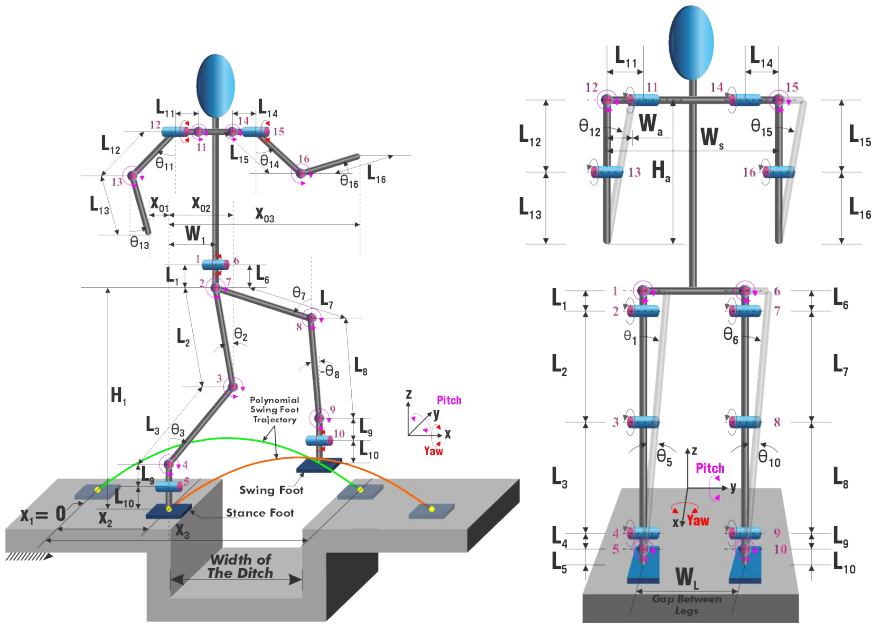
▪ **Wrist Trajectory (Sagittal View)**	
Cubic Polynomial, $z = \mu_0 + \mu_1 x + \mu_2 x^2 + \mu_3 x^3$	
x	**z**
x_{01}	h
$x_{01} + (d_w/2)$	$h + (d_w/8)$
$x_{01} + (3d_w/2)$	$h + (d_w/8)$
x_{03}	h

▪ **Hip Trajectory (Top View / Horizontal Plane)**	
Cubic Polynomial, $z = \mu_0 + \mu_1 x + \mu_2 x^2 + \mu_3 x^3$	
x	**z**
x_1	0
$x_1 + (d_w/2)$	$d_w/10$
$x_1 + (3d_w/2)$	$d_w/10$
x_3	0

▪ **Swing Foot Trajectory (Sagittal View)**					
Quadratic		**Cubic Polynomial**		**Fifth Order Polynomial**	
$z = \mu_1 x + \mu_2 x^2$		$z = \mu_0 + \mu_1 x + \mu_2 x^2 + \mu_3 x^3$		$z = \mu_0 + \mu_1 x + \mu_2 x^2 + \mu_3 x^3 + \mu_4 x^4 + \mu_5 x^5$	
x	**z**	**x**	**z**	**x**	**z**
x_1	0	x_1	0	x_1	0
$x_3 - (x_1/2)$	$d_w/2$	$x_1 + (d_w/2)$	$d_w/3$	$x_1 + (2d_w/5)$	$d_w/4$
x_3	0	$x_1 + (3d_w/2)$	$d_w/3$	$x_1 + (4d_w/5)$	$d_w/2$
		x_3	0	$x_1 + (6d_w/5)$	$d_w/2$
				$x_1 + (8d_w/5)$	$d_w/4$
				x_3	0

2.2 Forward & Inverse Kinematics of the Biped Robot

The main aim of this research is to generate a 3D smooth gait while crossing the ditch in both sagittal and frontal planes. Figure 2 shows the schematic diagram of the biped robot while crossing the ditch in both sagittal and frontal planes, respectively. To generate the gait cycle systematically a forward kinematic analysis is essential. Initially, the coordinate frames are assigned at each joint of the biped robot and Denavit-Hartenberg (D-H) notation is used for obtaining the position and orientation of the end effector. Table 2 shows the D-H parameters for two legs and two hands of AZAD-16.



(a) (b)

Figure 2
Schematic model of AZAD-16 demonstrating the various revolute joint angles in the (a) sagittal plane and (b) frontal plane

Table 3
DH parameters of AZAD-16

Frames	Link	Joint Angle (θ_i)	Twist Angle (α_i)	Link Length (a_i)	Joint Offset (b_i)
•	**Right Leg**				
0 to 1	1	θ_1	90	L_1	0
1 to 2	2	θ_2	0	L_2	0
2 to 3	3	θ_3	0	L_3	0
3 to 4	4	θ_4	-90	L_4	0
4 to 5	5	θ_5	0	L_5	0

•	**Left Leg**				
0 to 1	6	θ_6	90	L_6	0
1 to 2	7	θ_7	0	L_7	0
2 to 3	8	θ_8	0	L_8	0
3 to 4	9	θ_9	-90	L_9	0
4 to 5	10	θ_{10}	0	L_{10}	0
•	**Right Arm**				
0 to 1	11	θ_{11}	90	L_{11}	Y_1
1 to 2	12	θ_{12}	-90	L_{12}	0
2 to 3	13	θ_{13}	0	L_{13}	0
•	**Left Arm**				
0 to 1	14	θ_{14}	90	L_{14}	Y_1
1 to 2	15	θ_{15}	-90	L_{15}	0
2 to 3	16	θ_{16}	0	L_{16}	0

Once the polynomial trajectories are assigned for the foot and wrist in the sagittal plane and the hip in the horizontal plane, the gait generated from the various limbs of the biped robot is calculated from the concept of inverse kinematics. The initial joint angles of the upper and lower limbs of the swing leg (that is, θ_2 and θ_3) are obtained by using a closed form of inverse kinematic equations given in Eqn. (1) and Eqn. (2).

$$\theta_3 = sin^{-1}\left(\frac{H_1 L_2 \sin\vartheta + W_1(L_3 + L_2 cos\vartheta)}{(L_3 + L_2 cos\vartheta)^2 + (L_2 sin\vartheta)^2}\right) \tag{1}$$

$$\vartheta = cos^{-1}\left(\frac{{H_1}^2 + {W_1}^2 - {L_2}^2 - {L_3}^2}{2L_2L_3}\right) \tag{2}$$

Where H_1 is the height of the hip from the swing foot which can be calculated by using the relation $H_1 = L_3 cos\theta_3 + L_2 cos\theta_2$; and W_1 is the distance measured from the swing foot to the hip in 'X' direction which can be calculated from the following relation that is, $W_1 = L_3 sin\theta_3 + L_2 sin\theta_2$; Further, the joint angle θ_2 obtained from the following relation that is, $\theta_2 = \theta_3 - \vartheta$.

Similarly, the joint angles (that is, θ_{11} and θ_{13}) of the swing hand in sagittal plane is obtained by following Eqn. (3) and Eqn. (4).

$$\theta_{11} = cos^{-1}\left(\frac{{H_h}^2 + {W_h}^2 + {L_{12}}^2 - {L_{13}}^2}{2L_{12}\sqrt{{H_h}^2 + {W_h}^2}}\right) + \beta \tag{3}$$

$$\theta_{13} = sin^{-1}\left(W_h - \left(L_{12} sin\left(\frac{\theta_{11}}{\theta_{13}}\right)\right)\right) \tag{4}$$

Where H_h is height of the hand in wrist position to shoulder ($H_h = L_{12} cos\theta_{11} + L_{13} cos\theta_{13}$); and W_h is width of the hand in wrist to trunk ($W_h = L_{12} sin\theta_{11} + L_{13} sin\theta_{13}$). Moreover, the value β is calculated by using the following relation (that is, $\beta = tan^{-1}\left(\frac{W_h}{H_h}\right)$.

Further, the various joint angles (that is, $\theta_1, \theta_5, \theta_6, \theta_{10}, \theta_{12}, and\ \theta_{15}$) of the biped robot in frontal plane are given below.

$$\theta_1 = \theta_6 = tan^{-1}\left(\frac{W_L}{H_1}\right) \tag{5}$$

$$\theta_5 = \theta_{10} = tan^{-1}\left(\frac{W_L}{2H_2}\right) \tag{6}$$

$$\theta_{12} = \theta_{15} = tan^{-1}\left(\frac{W_a}{H_a}\right) \tag{7}$$

Where $H_2 = L_7 cos\theta_7 + L_8 cos\theta_8$, height of the arm $H_a = L_{15} cos\theta_{14} + L_{16} cos\theta_{16}$, $W_a = W_s/8$, W_s is distance between shoulder joints, and W_L is the distance between both legs.

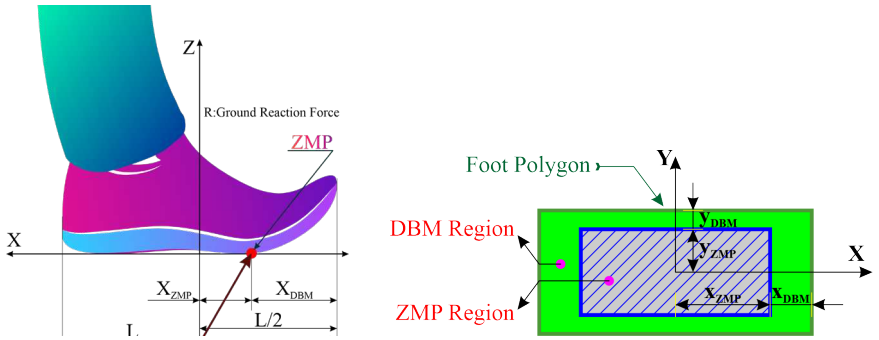
2.3 Dynamic Balance Margin

In the current study, the stability of the biped robot while crossing the ditch is obtained by the concept of ZMP. The ZMP in the x and y directions is calculated by using the following Eqn. (8) and Eqn. (9), respectively.

$$x_{ZMP} = \frac{\sum_{i=1}^{16}(-I_i\dot{\omega}_i + m_i x_i(\ddot{z}_i - g) + m_i\ddot{x}_i z_i)}{\sum_{i=1}^{16} m_i(\ddot{z}_i - g)} \tag{8}$$

$$y_{ZMP} = \frac{\sum_{i=1}^{16}(-I_i\dot{\omega}_i + m_i y_i(\ddot{z}_i - g) + m_i\ddot{y}_i z_i)}{\sum_{i=1}^{16} m_i(\ddot{z}_i - g)} \tag{9}$$

Where x_i, y_i, and z_i indicate the lumped mass coordinates, I_i represent the moment of inertia of the link in (kg-m^2), $\dot{\omega}_i$ denotes angular acceleration in (rad/s^2), m_i represents the mass of the link (kg), g indicates the acceleration due to gravity (m/s^2), and $\ddot{x}_i, \ddot{y}_i, and\ \ddot{z}_i$ represent acceleration in x, y and z direction for i^{th} link in (m/s^2).



(a) (b)

Figure 3
Schematic diagram showing (a) relation between ZMP and DBM, (b) top view of foot support showing ZMP region enclosed under DBM region

If the ZMP is falling nearer to the center of the foot, then the robot is more dynamically balanced. Suppose the ZMP does not fall inside the foot support polygon then move all the links and joints of the biped robot systematically towards the center of the foot. Therefore, the ZMP pushes inside the foot support polygon. Figure 3 shows the regions of ZMP and DBM in x and y directions along with ZMP. The DBM of the biped robot is calculated by following relations.

$$x_{DBM} = \frac{Length\ of\ the\ foot\ support}{2} - |x_{ZMP}| \tag{10}$$

$$y_{DBM} = \frac{Width\ of\ the\ foot\ support}{2} - |y_{ZMP}| \tag{11}$$

2.4 Dynamic Analysis of the Biped Robot

The dynamics of the 16-DOF biped robot is useful to estimate the torque required at each joint. In the current research work, the authors have considered the Lagrange Euler (L-E) formulation for calculating the dynamics. The torque required (τ_i) at each joint is calculated by the following equations.

$$\tau_i = \sum_{j=1}^{n} M_{ij}(\theta)\ddot{\theta}_j + \sum_{j=1}^{n}\sum_{k=1}^{n} h_{ijk}\dot{\theta}_j\dot{\theta}_k + G_i \tag{12}$$

Where $i, j = 1, 2, 3 \ldots\ldots\ldots\ldots 16;$ $\theta, \dot{\theta}_j\ and\ \ddot{\theta}_j$ are represented as angular displacement, angular velocity and angular acceleration at various joints. Further, the expanded terms of inertia forces (M_{ij}), centrifugal/Coriolis forces (h_{ijk}) and gravity forces (G_i) are provided below.

Inertia term, $M_{ij} = \sum_{p=\max(i,j)}^{n} Trace\left[d_{pj} I_p d_{pi}^T\right]$ (13)

Centrifugal/ Coriolis acceleration term,

$$h_{ijk} = \sum_{p=\max(i,j)}^{n} Trace\left[\frac{\partial(d_{pj})}{\partial\theta_k} I_p d_{pi}^T\right] \tag{14}$$

Gravity term, $G_i = -\sum_{p=i}^{n} m_p g d_{pi}\, {}^{p}\bar{r}_p$ (15)

Where I_p, ${}^{p}\bar{r}_p$ and g indicate the moment of inertia (kg-m/sec^2), the center of mass location (m) and acceleration due to gravity (m/ sec^2), respectively.

The amount of average power required for a 16-DOF biped robot to generate the gait while crossing the ditch is determined by the estimated torque and angular velocity of each joint. The equation required for calculating the average power is as follows.

Power Consumption, $P = \sum_{i=1}^{16} \int_{t_0}^{t_f} \left|\tau_i(t)\dot{\theta}_i(t)\right| dt$ (16)

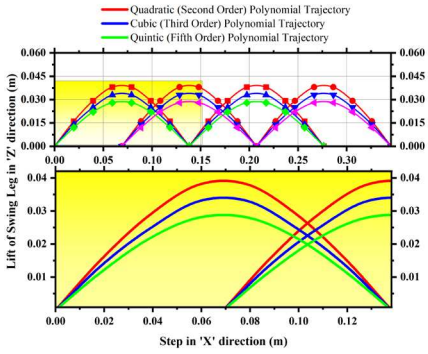
Here, t_0 and t_f represent the initial and final times, respectively. To achieve the desired angular displacement, it is necessary to determine the precise amount of joint acceleration that needs to be delivered to the actuator. The equation for joint acceleration ($\ddot{\theta}$) can be obtained by solving Eqn. 12, which is shown in Eqn. 17.

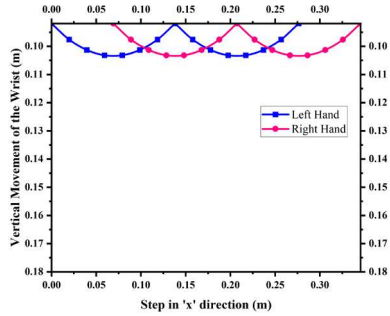
$$\ddot{\theta}_j = \tau_{i,actual} + \sum_{j=1}^{n} M_{ij}(\theta)^{-1}\left[-G_i - \sum_{j=1}^{n}\sum_{k=1}^{n} h_{ijk}\dot{\theta}_j\dot{\theta}_k\right] \quad (17)$$

Where $\tau_{i,actual} = \tau_i \sum_{j=1}^{n} M_{ij}(\theta)^{-1}$, represents the actual amount of torque required at each individual joint or servo actuator to rotate it by $\theta_{i,final} - \theta_{i,initial}$.

3 Results and Discussions

Once the mathematical model is developed the performance of the biped robot is measured in terms of dynamic balance margin while crossing the ditch in terms of computer simulations as well as the real biped robot that is, AZAD-16. The designed gait generation algorithm requires the initial position of the upper and lower links of the swing leg and hand. The necessary joint angles of the swing leg are $\theta_2 = 40^{\circ}$, and $\theta_3 = -30^{\circ}$; similarly, the initial joint angles of the; swing hand are $\theta_{10} = 40^{\circ}$ and $\theta_{13} = -40^{\circ}$ respectively. Initially, the DBM of the biped robot is tested in terms of swing foot trajectory. In this research, the authors assigned three varieties of swing foot trajectories which are derived from quintic, cubic and quadratic polynomial equations as shown in Figure 4 (a). The initial boundary condition of the swing foot was taken as, $x_1 = 0$ over a step length of 0.1382 m. The horizontal distance, or step size and time restriction for the gait cycle are constant for all situations. The wrist trajectory of both hands which were obtained from the cubic polynomial equations in the sagittal plane as shown in Figure 4 (b). Furthermore, Figures 4 (c) and 4 (d) depict the hip trajectory of the biped robot in the top or horizontal plane and sagittal plane which were derived as cubic polynomial and straight-line trajectories respectively. It is to be noted that the execution of the generated ditch crossing gait results in a straight-line horizontal hip trajectory or constant hip height throughout the gait which helps to maintain the dynamic balancing of the biped robot while performing the gait.





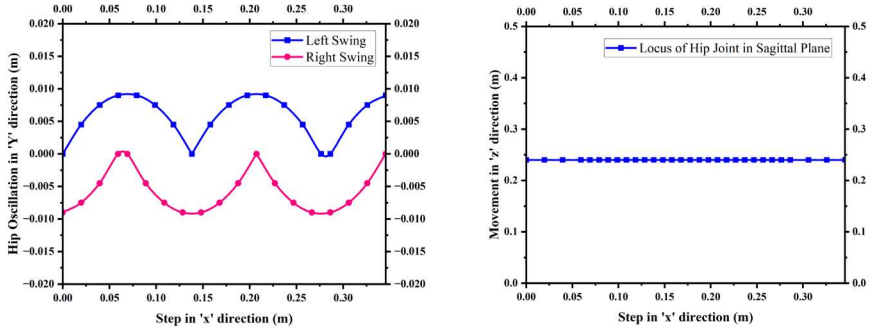


Figure 4

Graph showing (a) various foot trajectories in sagittal plane, (b) cubic polynomial trajectory of wrist end in sagittal plane, (c) cubic polynomial hip trajectory in top plane and (d) straight line hip trajectory in sagittal plane representing locus of hip height along 'x' direction

The results related to the swing foot trajectory are compared with the quadratic (second order) and quintic (fifth order) polynomial swing foot trajectories in terms of ZMP and DBM. Figures 5 (a) and (b) illustrate the variation of the ZMP measurement for all swing foot trajectories in the X and Y directions where the center of the stance foot is indicated by the zero value of the vertical axis. Further, Figures 6 (a) and (b) show the variation of DBM in X and Y direction. The investigation reveals that the cubic polynomial swing foot trajectory performs more dynamically balanced gaits when compared with quadratic and quintic polynomials. Because the ZMP is falling closer to the center of the foot in the case of cubic polynomial swing foot trajectory when compared with the quadratic and quintic polynomial trajectory. In addition, Figures 7 (a) and (b) depict the average DBM of the biped robot while crossing the ditch after assigning the quadratic, cubic and quintic polynomial trajectories in both X and Y directions. It has been observed that the cubic polynomial swing foot trajectory is more dynamically balanced when compared to quadratic and quintic polynomial trajectories.

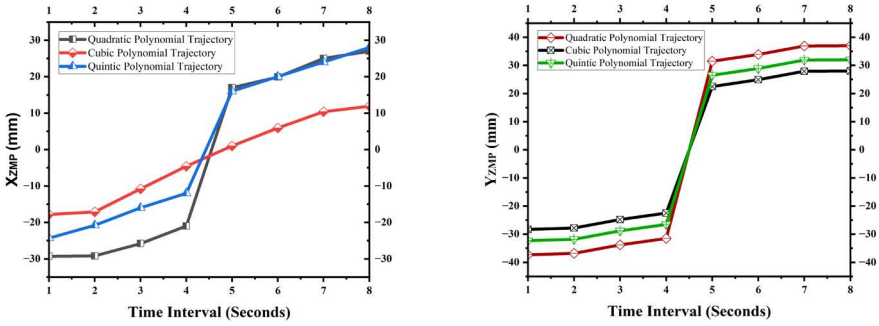


Figure 5

Variation of ZMP (a) X-direction and (b) Y- direction

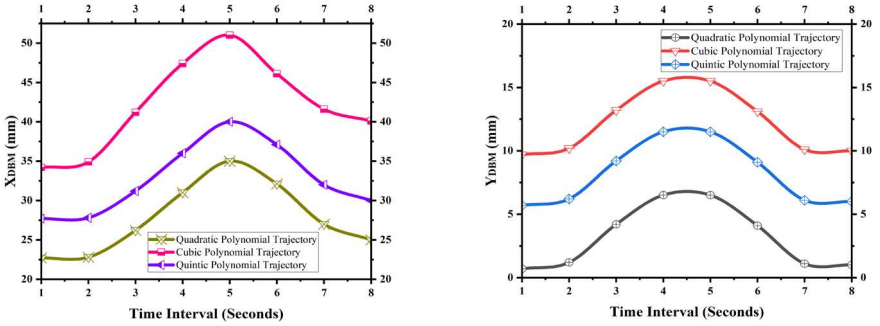


Figure 6
Variation of DBM (a) X-direction and (b) Y- direction

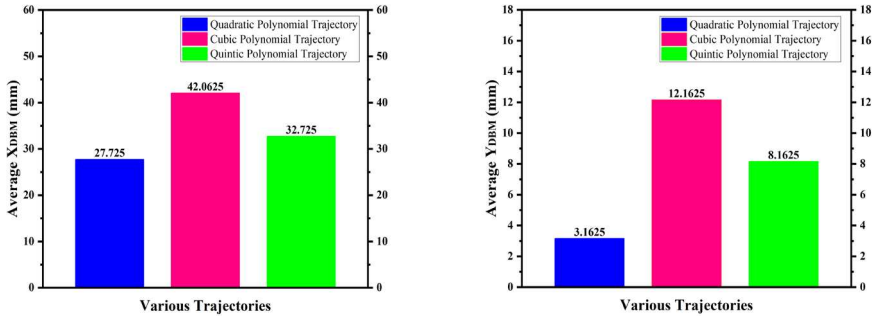


Figure 7
The bar chart shows the average DBM (a) in the x direction and (b) in the y direction as a consequence of various polynomial trajectories of the swing foot

Figure 8 (a) shows the variation of joint angles obtained from various joints of the biped robot only for the case of cubic polynomial swing foot trajectory while crossing the ditch in both SSP and DSP. The variation in the joint angles θ_2, θ_3, θ_4, θ_7, θ_8, θ_9, θ_{11}, θ_{13}, and θ_{16} causes the pitch motion of servo actuators and variation in the joint angles θ_1, θ_5, θ_6, θ_{10}, θ_{12}, and θ_{15} causes yaw motion in servo actuators for both SSP and DSP cases. It has been observed that the variation of joint angles in SSP is high. Whereas in DSP, the variation of joint angles is very less due to short time intervals of time. It has also been observed that the hip joints 2 and 6 are producing the pitch motion that is necessary to generate the ditch crossing gait and is showing a huge variation as depicted in Figure 8 (a), the joints 3–8, 4–9, 11–14, and 13–16 are obtaining the similar trend. Similarly, Figure 8 (b) shows the variation of angular velocity for all joints of the biped robot while crossing the ditch. It has been observed that the corresponding revolute joints 11 and 14 demonstrate a significantly higher variation in the angular velocity (rad/s) followed by joints 13 and 16 in DSP. DSP phases. The angular velocity changes from positive to negative which indicates the acceleration and deceleration within the gait range during the DSP phase.

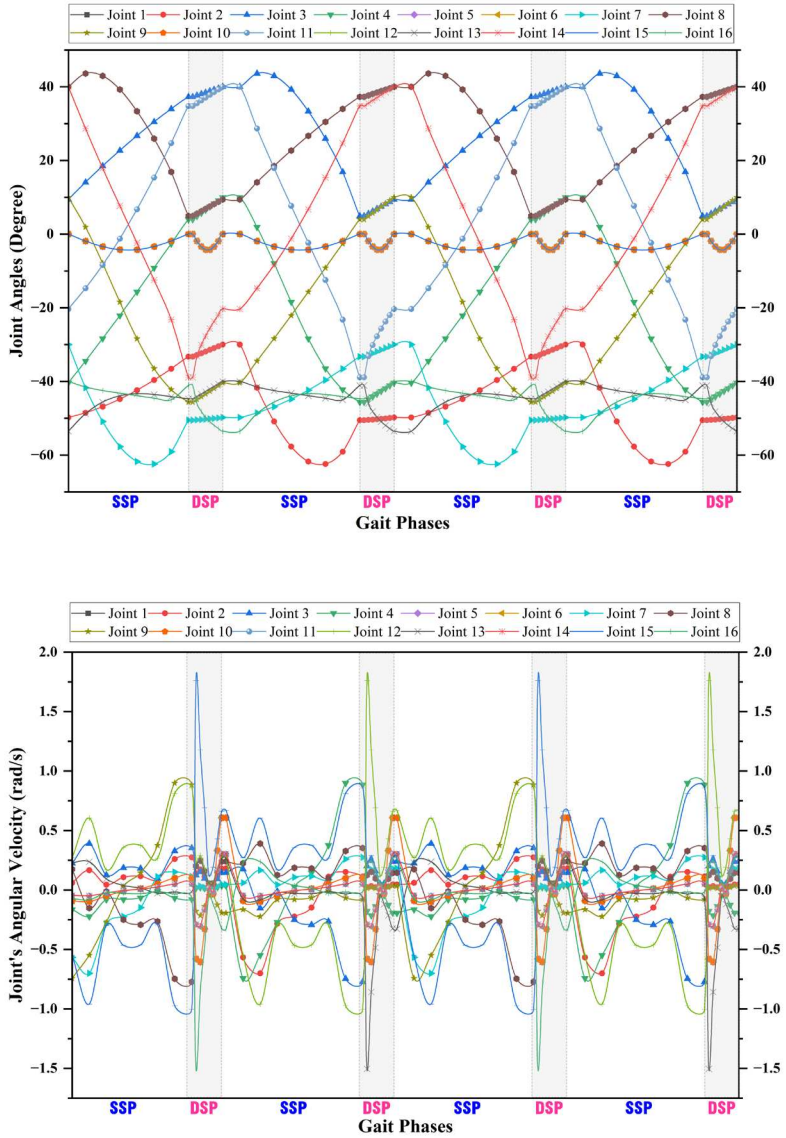
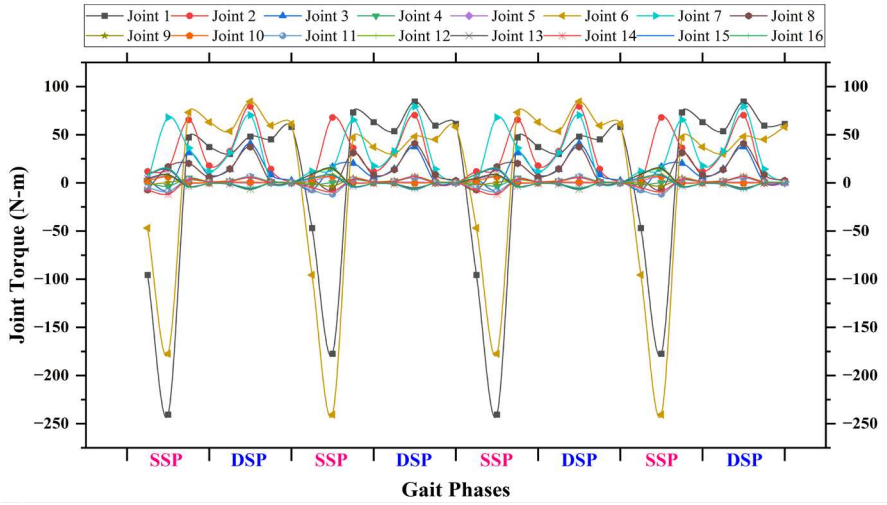


Figure 8

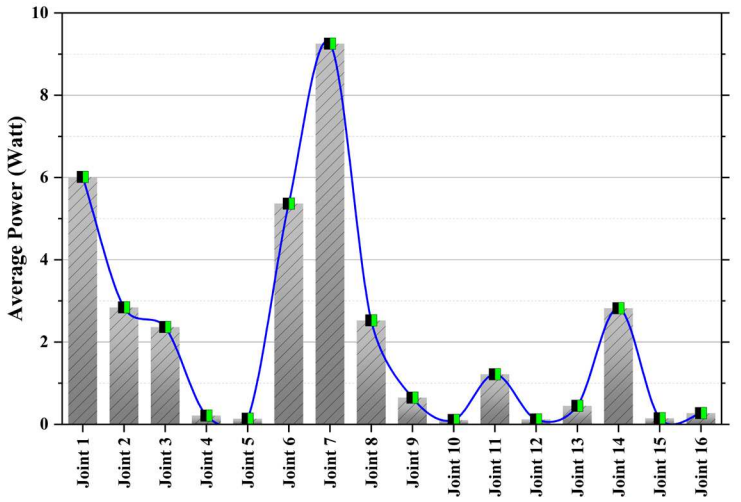
Variation of the joint parameters in SSP compared with its variation in DSP. (a) joint angles and (b) joint angular velocity



Schematic diagram showing the variation of torque required at various joints of the biped robot.

The cyclic variations in the magnitude of the actuator torque produced at each joint of the biped robot at various gait phases are shown in Figure 9. From Figure 9, it can be seen that the torque required at the hip joint is more than other joints. The magnitude of required joint torque is highest for yaw (joints 1 and 6) and pitch (joints 2 and 7) motions for both swing and stance legs. Because while exchanging the leg support, the hip joint of the leg carries the weight of the lower links and joints of the swing leg. It has been found that the hip joint of the stance leg consumes more torque when compared to the hip joint of the swing leg. Because the hip joint of the stance leg supports the whole body without falling on the ground. Further, figure 10 shows the average power consumption at various joints of the biped robot while crossing the ditch. It is important to note that, the power consumption is the product of individual joint torque and angular velocity. Despite, the highest torque obtained at joints 1 and 6, it has been observed that the power consumption is comparatively smaller than joint 7 due to lower angular velocity and the least variation of the joint angles throughout the gait cycle. Whereas joint 7 experiences the higher variation of joint angles and consequently consumes the highest power individually.

Figure 11 shows the simulation results of the 16-DOF biped robot in both sagittal and frontal plane. The investigation shows that the swing foot of the biped robot while crossing the ditch is following the cubic polynomial trajectory. It has also been observed that all the links and joint angles make proper gait without any fail and generate dynamically balanced gaits while crossing the ditch in both sagittal and frontal planes.



Schematic diagram showing the variation of average power consumption at different joints

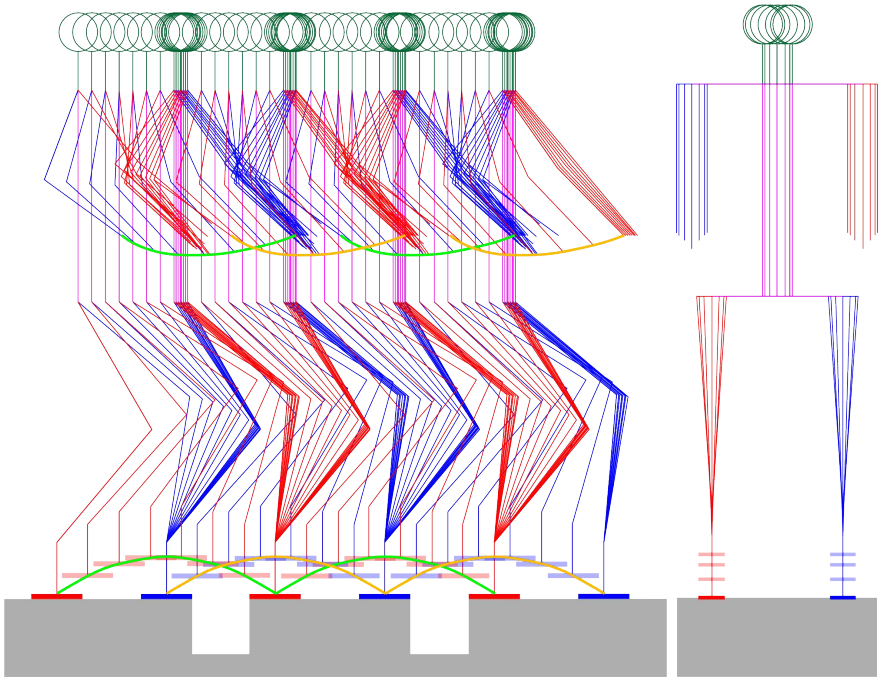
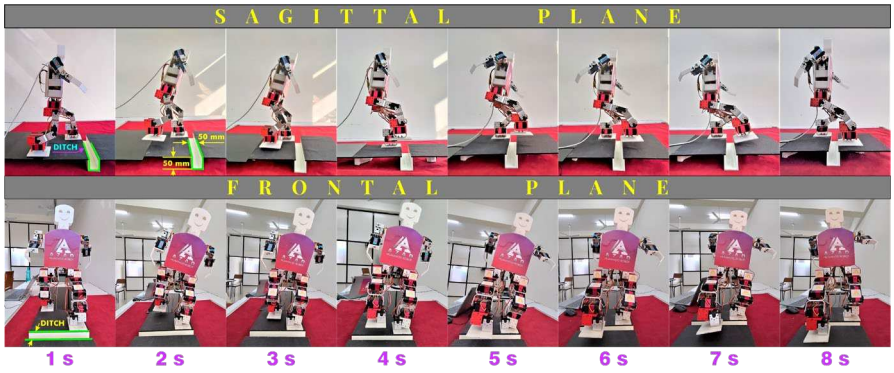


Figure 11
Stick diagram showing simulation of 16-DOF biped robot while crossing the ditch (a) sagittal plane and (b) frontal plane

Finally, the obtained gait angles from the simulations are fed into the real biped robot (that is, AZAD-16) developed in the Robotics Lab at MANIT Bhopal is shown in Figure 12. From Figure 12 it can be seen that the AZAD-16 biped robot generates a dynamically balanced gait while crossing the ditch in real time.



AZAD-16 performing the ditch crossing gait in the sagittal plane and frontal plane

Conclusion

In the present investigation, the authors successfully attempted to generate a gait for crossing ditch in both sagittal and frontal planes. Initially, the polynomials such as quadratic (second order), cubic (third order) and quintic (fifth order) are assigned for the swing foot while crossing the ditch. The result shows the cubic polynomial of the swing foot trajectory performing a more dynamically balanced gait when compared to quadratic and quintic polynomial trajectories. The concept of inverse kinematics has been adopted for obtaining the various joint angles. The dynamics of the 16-DOF biped robot is calculated by using the Lagrange-Euler formulation which helped in determining the torque at each joint of the biped robot. The hip joints of the swing and stance foot required more torque while crossing the ditch when compared to other joints of both the swing and stance leg. Moreover, research also reveals that the torque required at the hip joint of the swing leg is high as compared with the stance leg. Further, the knee joint of the biped robot consumes less torque than the hip joint but generates the highest torque among the rest of the other joints. A simulation study has been conducted in MATLAB. Finally, the obtained joint angles are fed into the real biped robot that is, AZAD-16 and verified the simulation study with the real biped robot walking in terms of dynamic balance margin (DBM).

References

[1] M. Vukobratovic, A. A. Frank, and D. Juricic, “On the Stability of Biped Locomotion,” *IEEE Trans. Biomed. Eng.*, vol. BME-17, no. 1, pp. 25–36, Jun. 1970, doi: 10.1109/TBME.1970.4502681.

[2] C. K. Chow and D. H. Jacobson, “Further studies of human locomotion:

Postural stability and control," *Math. Biosci.*, vol. 15, no. 1–2, pp. 93–108, Jun. 1972, doi: 10.1016/0025-5564(72)90065-X.

[3] M. A. Townsend and T. C. Tsai, "Biomechanics and modelling of bipedal climbing and descending," *J. Biomech.*, vol. 9, no. 4, pp. 227–239, Jun. 1976, doi: 10.1016/0021-9290(76)90008-7.

[4] R. Katoh and M. Mori, "Control method of biped locomotion giving asymptotic stability of trajectory," *Automatica*, vol. 20, no. 4, pp. 405–414, Jun. 1984, doi: 10.1016/0005-1098(84)90099-2.

[5] R. Katoh and M. Mori, "Control method of biped locomotion giving asymptotic stability of trajectory," *Automatica*, vol. 20, no. 1, pp. 229–234, Jun. 1982, doi: 10.1016/0005-1098(84)90099-2.

[6] T. Mita, T. Yamaguchi, T. Kashiwase, and T. Kawase, "Realization of a high speed biped using modern control theory," *Int. J. Control*, vol. 40, no. 1, pp. 107–119, Jun. 1984, doi: 10.1080/00207178408933260.

[7] Y. Hürmüzlü and G. D. Moskowitz, "The role of impact in the stability of bipedal locomotion," *Dyn. Stab. Syst.*, vol. 1, no. 3, pp. 217–234, Jun. 1986, doi: 10.1080/02681118608806015.

[8] A. Takanishi, M. Tochizawa, T. Takeya, H. Karaki, and I. Kato, "Realization of Dynamic Biped Walking Stabilized with Trunk Motion Under Known External Force," in *Advanced Robotics: 1989*, K. J. Waldron, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989, pp. 299–310. [Online]. Available: http://link.springer.com/10.1007/978-3-642-83957-3_21

[9] A. Takanishi, H. Lim, M. Tsuda, and I. Kato, "Realization of dynamic biped walking stabilized by trunk motion on a sagittally uneven surface," in *IEEE International Workshop on Intelligent Robots and Systems, Towards a New Frontier of Applications*, Jun. 1990, pp. 323–330. doi: 10.1109/IROS.1990.262408.

[10] Y. F. Zheng, "A neural gait synthesizer for autonomous biped robots," in *IEEE International Workshop on Intelligent Robots and Systems, Towards a New Frontier of Applications*, Jun. 1990, pp. 601–608. doi: 10.1109/IROS.1990.262457.

[11] "Real-time neural network control of a biped walking robot," *IEEE Control Syst.*, vol. 14, no. 1, pp. 41–48, Jun. 1994, doi: 10.1109/37.257893.

[12] T. Arakawa and T. Fukuda, "Natural motion trajectory generation of biped locomotion robot using genetic algorithm through energy optimization," in *IEEE International Conference on Systems, Man and Cybernetics*, Jun. 1996, vol. 2, pp. 1495–1500. doi: 10.1109/ICSMC.1996.571368.

[13] L. Magdalena and F. Monasterio-Huelin, "A Fuzzy logic controller with learning through the evolution of its knowledge base," *Int. J. Approx. Reason.*, vol. 16, no. 3–4, pp. 335–358, Jun. 1997, doi: 10.1016/S0888-613X(97)80098-9.

[14] G. Abba and N. Chaillet, “Robot Dynamic Modeling Using a Power Flow Approach with Application to Biped Locomotion,” *Auton. Robots*, vol. 6, no. 1, pp. 39–52, Jun. 1999, doi: 10.1023/A:1008820525412.

[15] Y. Nakamura *et al.*, “Humanoid robot simulator for the METI HRP Project,” *Rob. Auton. Syst.*, vol. 37, no. 2–3, pp. 101–114, Jun. 2001, doi: 10.1016/S0921-8890(01)00152-X.

[16] J. Chestnutt, J. Kuffner, K. Nishiwaki, and S. Kagami, “Planning biped navigation strategies in complex environments,” 2003.

[17] C. Sabourin and O. Bruneau, “Robustness of the dynamic walk of a biped robot subjected to disturbing external forces by using CMAC neural networks,” *Rob. Auton. Syst.*, vol. 51, no. 2–3, pp. 81–99, Jun. 2005, doi: 10.1016/j.robot.2005.02.001.

[18] J. R. T. Puga, F. M. T. Silva, and V. M. F. Santos, “MOTION PLANNING AND CONTROL STRATEGIES FOR A DISTRIBUTED ARCHITECTURE HUMANOID ROBOT,” *IFAC Proc. Vol.*, vol. 39, no. 15, pp. 773–778, Jun. 2006, doi: 10.3182/20060906-3-IT-2910.00129.

[19] R. Ghorbani, Q. Wu, and G. G. Wang, “Nearly optimal neural network stabilization of bipedal standing using genetic algorithm,” *Eng. Appl. Artif. Intell.*, vol. 20, no. 4, pp. 473–480, Jun. 2007, doi: 10.1016/j.engappai.2006.09.007.

[20] P. R. Vundavilli and D. K. Pratihar, “Inverse dynamics learned gait planner for a two-legged robot moving on uneven terrains using neural networks,” *Int. J. Adv. Intell. Paradig.*, vol. 1, no. 1, pp. 80–109, 2008.

[21] P. R. Vundavilli and D. K. Pratihar, “Dynamically balanced optimal gaits of a ditch-crossing biped robot,” *Rob. Auton. Syst.*, vol. 58, no. 4, pp. 349–361, Jun. 2010, doi: 10.1016/j.robot.2009.10.004.

[22] P. R. Vundavilli and D. K. Pratihar, “Gait Planning of Biped Robots Using Soft Computing: An Attempt to Incorporate Intelligence BT - Intelligent Autonomous Systems: Foundations and Applications,” D. K. Pratihar and L. C. Jain, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 57–85. doi: 10.1007/978-3-642-11676-6_4.

[23] A. Fattah and A. Fakhari, “Trajectory planning of walking with different step lengths of a seven-link biped robot,” in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2010, vol. 44106, pp. 1361–1369.

[24] A. P. Sudheer, R. Vijayakumar, and K. P. Mohandas, “Optimum stable gait planning for an 8 link biped robot using simulated annealing,” *Int. J. Simul. Model.*, vol. 10, no. 4, pp. 177–190, 2011.

[25] N. Kalamian and M. Farrokhi, “Dynamic walking of biped robots with obstacles using predictive controller,” in *2011 1st International eConference on Computer and Knowledge Engineering (ICCKE)*, 2011, pp. 105–110. doi: 10.1109/ICCKE.2011.6413334.

[26] L. S. R. Lathan, B. N. S. Rani, and P. R. Vundavilli, "Analytical approach for generating dynamically balanced gaits for obstacle crossing biped robot," in *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012)*, 2012, pp. 187–191.

[27] M. R. Kumar, L. S. Lathan, and P. R. Vundavilli, "Dynamically balanced obstacle crossing gait generation of a biped robot using neural networks," *Int. J. Mech. Robot. Syst.*, vol. 2, no. 3–4, pp. 232–253, 2015.

[28] R. K. Mandava and P. R. Vundavilli, "Whole body motion generation of 18-DOF biped robot on flat surface during SSP & DSP," *Int. J. Model. Identif. Control*, vol. 29, no. 3, pp. 266–277, 2018.

[29] R. K. Mandava and P. R. Vundavilli, "An analytical approach for generating balanced gaits of a biped robot on stairs and sloping surfaces," *Int. J. Model. Identif. Control*, vol. 33, no. 1, pp. 28–50, 2019.

[30] R. K. Mandava, K. Mrudul, and P. R. Vundavilli, "Dynamic motion planning algorithm for a biped robot using fast marching method hybridized with regression search," *Acta Polytech. Hung*, vol. 16, pp. 189–208, 2019.

[31] K. Bai, Y. Luo, G. Jiang, G. Jiang, and L. Guo, "High torque realization of the stepping over gait for a humanoid robot," *Ind. Robot Int. J. Robot. Res. Appl.*, vol. 47, no. 4, pp. 473–487, Jan. 2020, doi: 10.1108/IR-10-2019-0206.

[32] H. P. H. Anh and T. T. Huan, "Optimal walking gait generator for biped robot using modified Jaya optimization technique," *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, p. 382, 2020.

[33] M. Tsuru, A. Escande, A. Tanguy, K. Chappellet, and K. Harad, "Online Object Searching by a Humanoid Robot in an Unknown Environment," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2862–2869, Jun. 2021, doi: 10.1109/LRA.2021.3061383.

[34] L. Yang, G. Lai, Y. Chen, and Z. Guo, "Online Control for Biped Robot with Incremental Learning Mechanism," *Appl. Sci.*, vol. 11, no. 18, p. 8599, 2021.

[35] A. K. Kashyap, D. Parhi, and A. Pandey, "Improved Modified Chaotic Invasive Weed Optimization Approach to Solve Multi-Target Assignment for Humanoid Robot," *J. Robot. Control*, vol. 2, no. 3, Jun. 2021, doi: 10.18196/jrc.2377.

[36] A. K. Kashyap and D. R. Parhi, "Optimization of stability of humanoid robot NAO using ant colony optimization tuned MPC controller for uneven path," *Soft Comput.*, vol. 25, no. 7, pp. 5131–5150, Jun. 2021, doi: 10.1007/s00500-020-05515-1.

[37] V. Janardhan and R. Prasanth Kumar, "Online trajectory generation for wide ditch crossing of biped robots using control constraints," *Rob. Auton. Syst.*, vol. 97, pp. 61–82, Jun. 2017, doi: 10.1016/j.robot.2017.07.014.

[38] G. Gupta and A. Dutta, "Trajectory generation and step planning of a 12 DoF biped robot on uneven surface," *Robotica*, vol. 36, no. 7, pp. 945–970, 2018